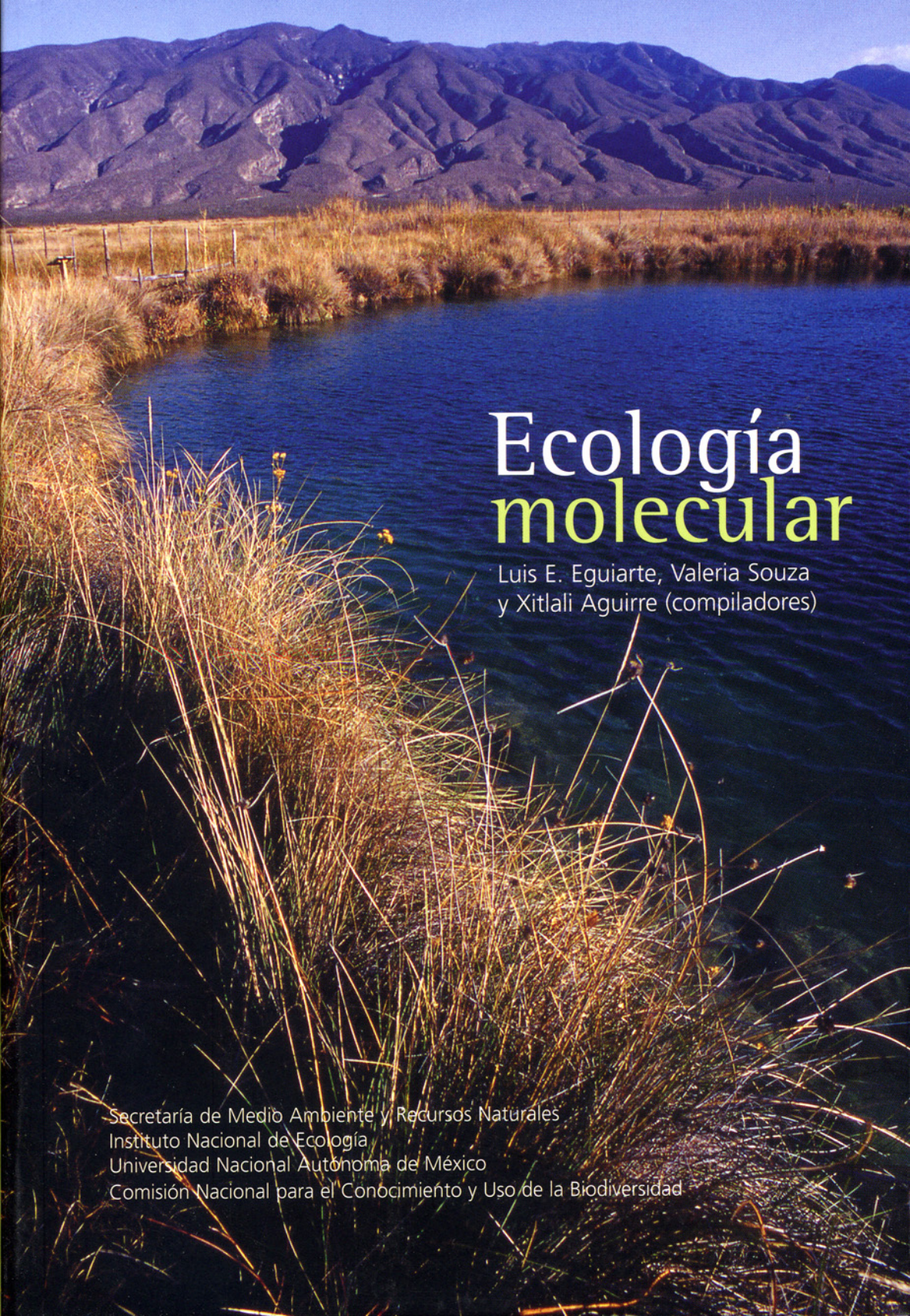


Ecología molecular

Luis E. Eguiarte, Valeria Souza
y Xitlali Aguirre (compiladores)

Secretaría de Medio Ambiente y Recursos Naturales
Instituto Nacional de Ecología
Universidad Nacional Autónoma de México
Comisión Nacional para el Conocimiento y Uso de la Biodiversidad

VISIT…

Ecología molecular

Ecología molecular

Luis E. Eguiarte, Valeria Souza
y Xitlali Aguirre
(Compiladores)

Secretaría de Medio Ambiente y Recursos Naturales
Instituto Nacional de Ecología
Universidad Nacional Autónoma de México
Comisión Nacional para el Conocimiento
y Uso de la Biodiversidad

Primera edición: abril de 2007

D.R. © Instituto Nacional de Ecología, Semarnat
Periférico sur 500, col. Insurgentes Cuicuilco, 04530
Delegación Coyoacán, México, D.F.
www.ine.gob.mx

Comisión Nacional para el Conocimiento
y Uso de la Biodiversidad (Conabio)
Ave. Liga Periférico-Insurgentes Sur 4903,
col. Parques del Pedregal, 14010
Delegación Tlalpan, México, D.F.
www.conabio.gob.mx

Coordinación editorial, diseño de interiores
y tipografía: Raúl Marcó del Pont Lalli
Diseño de la portada: Álvaro Figueroa
Foto de la portada: Claudio Contreras Koob

ISBN 978-968-817-839-3
ISBN 968-817-839-X

Impreso y hecho en México

Índice

Introducción a la Ecología Molecular 1
Luis E. Eguiarte y Valeria Souza

Agradecimientos 7

Primera parte. Teoría 9

Capítulo 1. La selección natural a nivel molecular 11
Amanda Castillo Cobián
Tipos de selección 12
The neutral expectation 13
Adaptación a nivel molecular 14
La selección a nivel genómico 37
Bibliografía 43

Capítulo 2. Flujo génico: métodos para estimarlo y marcadores moleculares 49
Erika Aguirre Planter
Métodos directos 50
Métodos indirectos 51

Métodos genealógicos 56
Bibliografía 59

Capítulo 3. Tamaño efectivo de la población 63
Alejandra Moreno Letelier
¿Qué significa el tamaño efectivo? 64
Métodos de estimación directos 65
Métodos ecológicos y demográficos 66
Métodos indirectos 71
Métodos de máxima verosimilitud y coalescencia 73
Aplicaciones 81
Conclusión 83
Bibliografía 84

Capítulo 4. Reconstrucción de la historia de cambio de los caracteres 87
León Martínez Castilla
Árboles, genes y organismos 89
Homología y alineación 104
Modelos de sustitución 112
Secuencias no alineables y métodos que no requieren alineación 123
Métodos de estimación de filogenias 127
Evaluación estadística de árboles 145
Métodos que pueden colocar OTUs contemporáneos en nodos internos y otras novedades 148
Bibliografía 152

Segunda parte. Temas fundamentales de la ecología molecular 161

Capítulo 5. Los análisis de paternidad, ¿para qué nos sirven? 163
Jordan Golubov y Jorge Ortega
La teoría y los métodos 165
Aplicación en animales 171
Aplicación en plantas 173
Conclusión 174
Bibliografía 178

Capítulo 6. La estimación de la endogamia y la relación entre la tasa de fecundación cruzada y los sistemas reproductivos en plantas con flores: una interpretación de su significado evolutivo 183
César A. Abarca García y Adriana López Villalobos
Las distintas formas de medir la endogamia 185
Estimaciones de la tasa de fecundación cruzada en poblaciones naturales de angiospermas 193
Bibliografía 204

Capítulo 7. La clonalidad y sus efectos en la biología de poblaciones 215
María C. Mandujano Sánchez
Organización jerárquica en plantas clonales y definiciones operativas 217
Tipos de clonación y frecuencia de clonalidad en las plantas 220
Ventajas y desventajas de la clonalidad 223
Demografía: estructura y dinámica 224
Distribución espacial de individuos en especies clonales y frecuencia de reclutamiento clonal en las poblaciones 228
Adecuación en plantas clonales 230
Estructura y variación genética en especies de plantas clonales 234
Análisis demográfico-genético 240
Discusión y consideraciones finales 242
Bibliografía 244

Capítulo 8. Ecología molecular de la conservación 251
Martha Rocha y Jaime Gasca
Un poco de historia 252
Pero, ¿qué relación guardan la genética y la conservación? 253
Especies en peligro, ¿a quién hay que proteger? 254
Principios básicos en genética de la conservación 254
Los enfoques actuales de la genética de la conservación 266
Conclusiones y perspectivas 271
Bibliografía 272

Tercera parte. La ecología molecular de los microorganismos 279

Capítulo 9. La recombinación: relevancia evolutiva y métodos de estimación, con énfasis en microorganismos 281
Rodolfo Salas Lizana
Mecanismos de la recombinación en eucariontes 282
Mecanismos de la recombinación en procariontes 284
Consecuencias de la recombinación 286
Las poblaciones microbianas: clonalidad vs. panmixia 288
Ejemplos de recombinación en bacterias 289
Ejemplos de recombinación en hongos filamentosos 291
Detectores y estimadores de la recombinación 292
Estimación de la recombinación 305
Consideraciones finales 310
Bibliografía 311

Capítulo 10. La especie como unidad evolutiva: uso de marcadores moleculares para su reconocimiento y delimitación, con especial énfasis en microorganismos 325
René Cerritos Flores
De lo natural a lo artificial y de lo universal a lo particular 326
Limitaciones de cada método para reconocer y delimitar unidades evolutivas 331
El concepto de especie en microorganismos, un problema aún no resuelto 337
Consideraciones finales 342
Bibliografía 343

Capítulo 11. Ecología evolutiva de bacterias y el concepto de especie: el caso de los rizobios 351
Claudia Silva y Pablo Vinuesa
La interacción rizobios-leguminosas 351
Genética de poblaciones y evolución en bacterias 355
Tres estudios de caso con rizobios 368
Conclusiones 377
Bibliografía 380

Capítulo 12. Ecología molecular en el estudio de comunidades bacterianas 393
Ana E. Escalante
El problema: comunidades y diversidad bacteriana, 394
microorganismos cultivables y no cultivables 394
¿Cómo mejorar la recuperación en cultivo? 396
¿Qué hacer con los no cultivables? Métodos moleculares y otros métodos 397
Métodos estadísticos para el análisis de las comunidades bacterianas 408
Revisión de estudios empíricos, crítica y perspectivas 414
Bibliografia 416

Cuarta parte. La ecología molecular de plantas y animales 425

Capítulo 13. Ecología evolutiva de las zonas de hibridación 427
Antonio González Rodríguez
Aspectos teóricos 428
Estudios de zonas de hibridación en plantas y animales 432
Bibliografía 435

Capítulo 14. Filogeografía y vertebrados 441
Ella Vázquez Domínguez
Algunos aspectos generales sobre filogeografía 442
La molécula estrella, el ADN mitocondrial 444
Teoría de coalescencia y métodos de análisis 447
Filogeografía de vertebrados 449
Filogeografía intraespecífica 453
Filogeografía comparada 459
Bibliografía 461

Capítulo 15. Filogeografía de aves mexicanas 467
Blanca Estela Hernández-Baños, Magali Honey-Escandón, María N. Cortés-Rodríguez y Gabriela García Deras
Métodos y análisis de la filogeografía 470
Estudios mexicanos 479
Bibliografía 490

Quinta parte. Las herramientas moleculares 497

Capítulo 16. Extracción de ácidos nucleicos 499
Luisa I. Falcón y Aldo Valera
El ADN en el laboratorio 500
ADN 500
Protocolos de extracción de ADN implementados en el laboratorio de Evolución Molecular y Experimental del Instituto de Ecología 502
Extracción de ADN de bacterias gram negativas 503
Extracción de ADN en plantas 506
Extracción de ADN en animales 511
ARN 512
Métodos de extracción del ARN 513
Bibliografía 515

Capítulo 17. Guía práctica sobre la técnica de PCR 517
Laura Espinosa Asuar
Tipos de técnicas 520
¿Qué se necesita para hacer un PCR y cómo conseguir estos reactivos en México? 523
Haciendo PCR: cómo hacerlo y algunos tips para que salga mejor 524
Métodos para visualizar el PCR 532
Recetas 536
Bibliografía 536

Capítulo 18. Breve revisión de los marcadores moleculares 541
Miroslava Rentaría Alcántara
Variación morfológica y variación molecular 541
Marcadores moleculares 542
Isoenzimas/aloenzimas 544
RAPDs 545
Microsatélites 548
ISSRs 550
RFLPs 552
AFLPs 554

Secuenciación de ADN 557
Microarrays (microarreglos) 559
Bibliografía 562

Capítulo 19. Inter Simple Sequence Repeats (ISSRs) 567
Andrea González y Xitlali Aguirre
Ventajas y desventajas 569
Aplicaciones 569
Bibliografía 571

Índice analítico 573

Introducción a la Ecología Molecular

Luis E. Eguiarte y Valeria Souza

La Ecología Molecular es una novedosa y vigorosa rama de la Ecología. De manera resumida la podemos definir como el empleo de herramientas moleculares para resolver problemas ecológicos. Tal aplicación ha abierto la puerta para el estudio de problemas que hace pocos años nos parecían insolubles aunque fascinantes y centrales. Al mismo tiempo ha generado nuevas preguntas y campos de conocimiento, tanto en lo ecológico como en lo evolutivo.

La Ecología Molecular se ha enfocado al uso de marcadores genéticos para resolver problemas poblacionales, como queda claro de hojear cualquier número de la revista *Molecular Ecology*. Y trata de resolver los problemas clásicos de la ecología de poblaciones: ¿cómo definir una población?; ¿dónde acaba y dónde inicia?; ¿cómo definir a las especies?; ¿cómo analizar a las poblaciones híbridas? y ¿cuáles son realmente las especies que les dieron origen?

Otros problemas clásicos dentro la Ecología Molecular han sido su uso para llevar a cabo análisis de paternidad, pregunta central en el estudio de la conducta, la evolución de los sistemas reproductivos y la selección sexual. Íntimamente relacionado con el análisis de paternidad está la estimación directa del flujo génico, y en este tenor, la Ecología Molecular rápidamente se confunde con la Genética de Poblaciones, disciplina que le da sustento a todos los análisis formales con marcadores moleculares.

Adicionalmente, la Ecología Molecular ha tratado de usar los marcadores genéticos para decidir si un organismos pertenece o no a una especie, y esta idea ha

sido llevada a niveles exquisitos en su uso en bacteriología, donde se ha revelado una diversidad insospechada de microorganismos, muchos de los cuales no se había presumido su existencia previamente (i.e., todo el dominio Archaea), ya que no se pueden cultivar con los métodos tradicionales de la microbiología.

Y se han abierto nuevas avenidas. Por el lado de la filogenia, la Ecología Molecular nos puede dar elementos para realizar estudios comparativos y entender, en un marco riguroso y formal, la evolución de las adaptaciones y la morfología. También se han usado con éxito ideas filogenéticas para definir prioridades de conservación. Recientemente, se ha desarrollado un nuevo campo que sugiere por fin una síntesis entre los estudios poblacionales y los estudios macroevolutivos, en un contexto ecológico. Éste es el estudio de las tasas de especiación, que junto con relojes moleculares nos permiten explorar mejor las radiaciones evolutivas, para tratar de contestar ¿por qué surgen más especies?, ¿por qué a esos ritmos, a esos tiempos?

También es muy importante mencionar un avance que consideramos crítico dentro de la Ecología Molecular: el estudio de los Metagenomas. La idea de Metagenoma es muy básica, de fuerza bruta: secuenciar el genoma completo de todos los organismos de un lugar. Lo que generalmente se hace es juntar una muestra de agua, sedimento o suelo, separar a los procariotes (ya que el genoma de los eucariotes es demasiado grande para ser secuenciado de esta forma) y de estos organismos secuenciar una muestra, más o menos aleatoria, de este total de ADN. Los detalles técnicos son complicados, pero la idea es que con este método se podría saber no sólo los genes, sino a partir de su análisis tener una idea de las funciones bioquímicas que realizan los organismos y así inferir como está funcionando el ecosistema. Es decir, a partir de los genomas hacer inferencias ecológicas a nivel ecosistema y fisiología (además de aportar datos sobre la diversidad, abundancia, filogenia y evolución de todos los microorganismos de una localidad).

En otras palabras, la Ecología Molecular incide a todos los niveles de análisis de la ecología moderna: a nivel de las poblaciones, para definirlas, para conocer su historia y sus parámetros evolutivos, paternidades, adecuaciones y flujo génico. A nivel evolutivo es claro que nos da datos para analizar a los organismos, poblaciones y comunidades en términos de Genética de Poblaciones y Evolución Molecular, y no ayuda en la estimación de los parámetros evolutivos clásicos (selección natural, deriva génica, mutación, migración, endogamia). A nivel fisiológico y funcional nos ayuda a descubrir las adaptaciones y vías bioquímicas. A nivel comunidades sirve para definir las especies que existen y sus densidades, particularmente en el caso de microorganismos, y a veces a detectar patrones coevolutivos y de interacción. Y

a nivel de los ecosistemas nos da elementos para comprender su funcionamiento, en especial en relación a más famosos caja negra, los microorganismos.

La Ecología Molecular, como una disciplina reconocida, es relativamente reciente. El nombre (aplicado a su concepción actual, por ejemplo, el Dr. Jaime Mora lo usó para su laboratorio, el Centro de Fijación del Nitrógeno, UNAM, con objetivos y metas diferentes a los que usamos actualmente) surge a finales de los años 1990. En particular, podemos citar el libro *Genes in Ecology* de la British Ecological Society, derivado de un simposio celebrado en 1991, y la aparición en 1992 de la revista *Molecular Ecology*. En 1998 G.R. Carvhalo editó *Advances in Molcular Ecology*, y recientemente (2004) apareció un primer libro de texto sobre el tema, *An introduction to Molecular Ecology,* de T. Beebee y G. Rowe, dirigido a estudiantes avanzados de licenciatura.

En particular en México, alrededor del Instituto de Ecología de la UNAM se ha estado formando un vigoroso grupo de jóvenes investigadores trabajando en el área de la Ecología Molecular, derivado de un sólido conjunto de investigadores interesados en el estudio de la Ecología de Poblaciones de plantas y animales, con una fuerte base cuantitativa y de genética de poblaciones. El inicio fue relativamente lento, por la falta de infraestructura física y de entrenamiento en Biología Molecular, pero recientemente el desarrollo ha sido impresionante y muy exitoso. Este libro es resultado de esta interacción y desarrollo.

Este libro está dirigido a personas con una formación biológica con énfasis en la ecología y evolución y con conocimientos básicos de genética general, estadística y biología molecular. La mayor parte de los capítulos, aunque comienzan "desde el principio", tiene un carácter avanzado, y su objetivo es guiar a los estudiantes de los últimos años de licenciatura, y principalmente a estudiantes graduados y a científicos interesados en iniciar investigaciones en los diferentes subtemas que componen el campo de la Ecología Evolutiva. Cada capítulo pretende ser una guía a los desarrollos e ideas más importantes, para que posteriormente el lector pueda adentrarse con más seguridad y claridad en estos campos. Los lectores que lo juzguen necesario pueden revisar conceptos básicos de evolución y genética de poblaciones en el texto que compilamos Juan Nuñez-Farfán y Luis E. Eguiarte (Evolución Biológica, 1999, UNAM, Conabio, México).

Esta obra está dividida en cinco partes. En la primera, denominada Teoría, revisamos en cuatro capítulos los aspectos teóricos fundamentales en los que se sustenta la Ecología Molecular, relacionados con la Genética de Poblaciones y con la Evolución Molecular. Iniciamos analizando, a nivel molecular, la fuerza evolutiva más importante para los ecólogos, la selección natural, que es el motor básico de la adaptación y de los procesos poblacionales. Amanda Castillo hace una actua-

lizada revisión del tema, describiendo los métodos estadísticos para detectarla a nivel molecular, ejemplifica los avances y termina describiendo los programas de computadora desarrollados recientemente para este tipo de análisis.

A continuación revisamos dos fuerzas evolutivas que se han estudiado con gran éxito recientemente usando marcadores moleculares. El primer trabajo, de Erika Aguirre, describe el flujo génico, tanto su estimación directa como indirecta, y algunos avances genealógicos y de coalescencia. Alejandra Moreno revisa las diferentes ideas, metodologías y fórmulas para estimar el tamaño efectivo y para entender el papel de la deriva génica, tocando diferentes métodos relacionados con la coalescencia y revisando las perspectivas para su estimación y estudio. La sección cierra con una revisión de las ideas y métodos de reconstrucción filogenética por León Martínez Castilla, donde se examinan los conceptos relacionados con árboles filogenéticos, y los diferentes métodos de reconstrucción filogenética.

El siguiente apartado contiene cuatro capítulos y trata de los algunos temas fundamentales de la Ecología Molecular. Inicia con el tema central dentro de esta disciplina, el estudio de la paternidad en las poblaciones naturales, donde Jordan Golubov y Jorge Ortega pasan revista a la teoría y ejemplos de estudios tanto en plantas como en animales. Posteriormente, se analizan dos problemas clásicos de la ecología molecular en plantas. Primero el de la endogamia y de su principal determinante en plantas, la tasa de fecundación cruzada en plantas, cuya autoría pertenece a César Abarca y Adriana López. En segundo lugar se encuentra el análisis de María del Carmen Mandujano de los niveles de clonalidad en angiospermas, quien inspecciona la literatura sobre el uso de marcadores moleculares, sus problemas evolutivos y técnicos y sus perspectivas.

Otra faceta central de la Ecología Molecular es un tema "sintético", que usa datos y métodos de diversas fuentes y tradiciones, pero que nos permite redondear ideas y conceptos tanto de Ecología Molecular como de Genética de Poblaciones. Nos referimos a la Genética de la Conservación, una de las aplicaciones potencialmente más importantes de la Ecología Molecular, que ha recibido recientemente considerable atención, tanto en libros de texto, simposios y revistas especializadas y de la cual Martha Rocha y Jaime Gazca presentan una interesante revisión actualizada.

Los métodos de la Ecología Molecular nos han ayudado a entender y definir mejor a las especies, poblaciones y en general a los proceso ecológicos y evolutivos en los microorganismos. La tercera parte de este libro contiene cuatro capítulos donde se usan métodos de ecología molecular para estudiar diferentes aspectos de estos organismos. En un primero trabajo, Rodolfo Salas revisa los avances en métodos para detectar la recombinación en microorganismos, tanto bacterias como

hongos. También discute los estudios empíricos y los programas disponibles para su estimación. En muchos microorganismos no se han observado directamente procesos o mecanismos sexuales o de recombinación, por lo que los marcadores genéticos y métodos estadísticos y de Genética de Poblaciones son la mejor, y a veces la única herramienta para medir la magnitud de esta recombinación. No debemos olvidar que estos métodos se pueden utilizar para estudiar los niveles de recombinación en otros genomas y organismos, tema que seguramente va a tener un desarrollo muy importante usando datos derivados de los proyectos genómicos y para analizar incidencias de enfermedades y susceptibilidades congénitas.

Como ya mencionamos, una aplicación clásica de la Ecología Molecular ha sido la de tratar de definir que es una especie, principalmente dentro de microorganismos, y derivar reglas para precisar los criterios de pertenencia de un organismo a una especie dada. Rene Cerritos desarrolla este tema considerando ejemplos tanto de bacterias como de hongos. El siguiente capítulo es, en parte, una aplicación de las ideas enunciadas en el anterior. En él Claudia Silva y Pablo Vinues analizan la Ecología Molecular del grupo de los rizobios, importantes simbiontes mutualistas de las leguminosas, y exploran su ecología, su diversidad, las especies y linajes que existen, así como su radiación evolutiva y biogeografía y echan mano de estas evidencias para adentrarse en el entendimiento de las especies de rizobios, ayudándonos a acercarnos a una mejor comprensión de lo que es una especie en bacterias, lo cual constituye uno de los problemas más importantes y paradójicos de la microbiología contemporánea.

Para concluir con el tópico de los microorganismos, Ana Escalante revisa los métodos e ideas usados para estudiar a las comunidades bacterianas con métodos de Ecología Molecular, otro de los temas clásicos de nuestra disciplina, que realmente ha abierto el estudio de los microorganismos a la ecología moderna a todos los niveles, especialmente al de las comunidades.

La cuarta parte de esta obra revisa tres estudios concretos donde se usan conceptos y métodos de Ecología Molecular en eucariontes para ejemplificar algunos de sus usos y aplicaciones y para ilustrar las preguntas que se trabajan, sus relaciones con otras disciplinas y sus limitaciones. Las zonas híbridas entre especies han sido consideradas un excelente laboratorio natural para entender el proceso de la especiación y analizar fenómenos evolutivos relacionados a la adaptación, selección natural y flujo génico. Antonio González describe los avances recientes en su estudio en plantas y animales con una perspectiva de Ecología Molecular.

A continuación se presentan dos revisiones sobre la filogeografía. Ella Vázquez repasa el tema de la filogeografía en general. Nos expone un poco de la historia de estas ideas y las características del genoma mitocondrial en animales, para

concluir discutiendo algunos ejemplos de estudios con mamíferos y otros vertebrados. Blanca Hernández, junto con sus colaboradores, nos ilustran sobre los patrones filogeográficos encontrados en diversos estudios de filogeografía con aves, especialmente en el territorio mexicano, discuten diferentes metodologías, como son el uso de las distribuciones de *mismatches* y el *Nested Clade Analysis*, y comparan datos de la literatura con sus propios resultados recientes.

La quinta sección trata de los métodos moleculares propiamente dichos. Miroslava Rentería, estudiante de doctorado, inicia con una revisión de los métodos utilizado. A continuación, Aldo Valera y Luisa Falcón, nos muestran con detalle el primer paso en (casi) todo estudio de Ecología Molecular, la extracción de ADN o ARN. Y discuten diferentes métodos de extracción del ADN y RNA, para varios grupos de organismos, con énfasis en bacterias y plantas. El siguiente capítulo, escrito por Laura Espinoza, trata de la "piedra angular" de la Ecología Molecular, la cual abrió las posibilidades ilimitadas del campo, la PCR (reacción encadenada de la polimerasa). Esta autora nos ofrece un serie de principios, de ideas y de consejos para poder llevar a cabo PCR exitosos. Por último, Andrea González y Xitlali Aguirre nos presentan el ISSR, un método muy sencillo que ha sido usado exitosamente para estudios en planta y animales, y que creemos que debe contar con una difusión mayor, dado lo sencillo, económico y reproducible que ha resultado.

Agradecimientos

Agrademos especialmente el apoyo de los proyectos SEP/CONACYT 44673-Q y 46475-Q, SEMARNAT/CONACYT CO10237 A1 y CO1-0246 A1 y PAPIIT/UNAM IN-223105 otorgados a Valeria Souza y/o a Luis E. Eguiarte, que nos permitieron realizar nuestro estudio sobre Ecología Molecular y avanzar en la edición del libro.

Los doctores Mark Olson del Instituto de Biología, UNAM y Arturo Becerra de la Facultad de Ciencias, UNAM, revisaron todo el manuscrito. También agradecemos a los revisores anónimos de la Conabio.

El libro se originó de un par de cursos que dimos en los posgrados de ciencias biológicas y biomédicas de la UNAM. Agradecemos a todos nuestro colegas y alumnos que dentro de los cursos, talleres y discusiones informales, alrededor de nuestro Laboratorio de Ecología Molecular y Experimental, participaron en las discusiones, estudios de campo y laboratorio y análisis y seminarios que dieron forma a este libro.

A Xitlali Aguirre inicialmente la reclutamos para que nos ayudara en las revisiones y correcciones del manuscrito y a perseguir a los coautores, pero su labor fue extraordinaria, por lo que aparece como coeditora del libro. Raúl Marcó del Pont Lalli, del INE, ayudó enormemente a la configuración final del libro. También queremos agradecer a todos los coautores que nos ayudaron a revisar los diferentes temas y capítulos.

La edición e impresión del libro gracias a un generoso apoyo del Instituto Nacional de Ecología, y agradecemos profundamente el interés y ayuda de la M. en C. Irene Pisanty. Para la edición del libro también contamos con apoyos parciales por parte de la Comisión Nacional para el Conocimiento y Uso de la Biodiversidad (Conabio) y del Instituto de Ecología de la Universidad Nacional Autónoma de México (UNAM).

Valeria y Luis quieren dedicar este libro a sus hijos, Luis Felipe y María Antonia, que han sufrido de manera razonablemente estoica la falta de atención de sus padres por culpa de la Ecología Molecular.

Primera parte

Teoría

Capítulo 1

La selección natural a nivel molecular

Amanda Castillo Cobián

Las dos ideas principales derivadas de la teoría Darwin-Wallace sobre la evolución de las especies señalan que los organismos son producto de una historia evolutiva, es decir, de ancestría-descendencia a partir de modificaciones de un ancestro en común. Además, que el mecanismo principal de la evolución es la selección natural de estas variaciones hereditarias (Futuyma, 1986). A partir de estas ideas se generaron dos campos de estudio dentro de la biología. Por un lado el estudio de la historia evolutiva de los organismos, y por el otro, la elucidación de las fuerzas evolutivas que moldean la biodiversidad y las adaptaciones que producen (Futuyma, 1986). El papel de la selección natural es ampliamente aceptado dentro de la comunidad científica, sobre todo en cuanto a caracteres morfológicos se refiere, es decir, a las variaciones fenotípicas, de las cuales se tienen abundantes ejemplos (Lack, 1947; Wiesenfeld, 1967; Bishop y Cook, 1975; Jope, 1976; Grant, 1986; Sato *et al.*, 1999; Zhang *et al.*, 2002). Sin embargo, la importancia de la selección natural a nivel molecular es materia de un continuo debate.

La selección se refiere a la reproducción diferencial de algunos fenotipos/genotipos sobre otros bajo ciertas condiciones ambientales que prevalecen en un momento determinado (Futuyma, 1986; Li, 1997). En otras palabras, es un mecanismo evolutivo que genera un cambio en las frecuencias relativas de los fenotipos/genotipos, de acuerdo a su adaptación relativa dentro de la población. Durante mucho tiempo se creyó que la selección era la principal

fuerza que mantenía la variación fenotípica y genotípica. Sin embargo, a partir del desarrollo de nuevas técnicas para el estudio de la variación a nivel ADN y proteínas se inició de manera formal el estudio de la evolución a nivel molecular (Zuckerkandl y Pauling, 1965). Los resultados de toda una serie de trabajos de genética de poblaciones con diferentes organismos arrojaron información acerca de la naturaleza de la variación a nivel molecular y trajeron como resultado la evidencia de que existía una gran variación genética contenida en las proteínas y genes, misma que no podía ser justificada del todo por efectos de la selección natural. Estas ideas culminan con la propuesta de la teoría neutra de la evolución (Kimura, 1968, 1983).

De acuerdo con la teoría neutra de la evolución molecular, la mayor parte de la variación al nivel molecular no es mantenida por la selección tipo darwiniano (selección adaptativa o positiva), sino resultado de un balance entre la deriva génica y las mutaciones neutras a casi neutras (Kimura, 1983; Ohta, 1992). Es decir, la mayor parte de la variación a nivel molecular ocurre de manera aleatoria y no tiene importancia adaptativa. En su conjunto, los postulados de la teoría neutra parecían regir en buena medida a la evolución molecular.

Debido a estas ideas, se ha generado un gran debate entre las discusiones de si existen pruebas para demostrar la presencia de la selección positiva a nivel molecular, y se han desarrollado numerosos métodos para su detección.

Tipos de selección

Básicamente, la selección natural a nivel molecular puede ser considerada de dos tipos: selección positiva (que incluye a la selección balanceadora o diversificadora, y a la adaptativa darwiniana) y selección negativa (purificadora; Li, 1997).

La selección positiva adaptativa es el mecanismo evolutivo mediante el cual nuevos mutantes poseen adecuaciones mayores que el promedio de la población, y las frecuencias de dichos mutantes se incrementan en la siguiente generación (Li, 1997). Por otra parte, en la selección negativa los nuevos mutantes poseen adecuaciones menores que el promedio de la población, y la frecuencia de estos mutantes disminuye en las siguientes generaciones (Li, 1997). La selección positiva balanceadora promueve la diversidad genética y la selección negativa disminuye o purga esta diversidad eliminando las variantes de la población.

El estudio de la adaptación a nivel molecular se ha abordado sólo de manera reciente como un intento de evaluar cuál es el papel de la selección a nivel

molecular, ya que inicialmente la mayor parte de los estudios parecían indicar que la selección adaptativa tendría un papel secundario (Suzuki y Gojobori, 1999). Pero ¿qué tal si los métodos desarrollados no eran lo suficientemente sensibles para detectar la selección de manera inequívoca y directa? ¿Cómo podemos aprovechar toda la información contenida en las bases moleculares para inferir el papel que la selección positiva juega dentro de la evolución molecular?

The neutral expectation

De manera histórica la selección positiva es estudiada indirectamente cuando un grupo de genes parece desviarse del modelo neutral, es decir, al no encajar en las predicciones neutrales se supone de forma indirecta que estas desviaciones son producto de la selección que las está modificando. Es entonces que para poder detectar la participación de la selección en una región del genoma o en un gen en particular, lo primero que se tiene que probar es el rechazo de la hipótesis nula de que la selección no ha actuado en las secuencias en cuestión. Los postulados de la teoría de Kimura (1968, 1983) nos permiten establecer qué patrones de evolución de las secuencias se deben esperar si existen desviaciones del modelo neutro originadas por la mutación y la deriva génica sin la participación de la selección.

Un resultado particularmente importante derivado del modelo neutro es que la tasa de cambio a la cual un nucleótido es remplazado (sustituido) por otro dentro de una población es igual a la tasa de mutación o cambio de ese sitio sin importar el tamaño poblacional. El número de mutaciones se encuentra en balance con la baja probabilidad de que cada mutación se fije en una población grande. Además, la teoría neutra hace predicciones acerca de los patrones de polimorfismo esperados dentro de las secuencias de las mismas especies. Por ejemplo, la probabilidad de que dos muestras tomadas al azar difieran en una posición particular del genoma se determina por la siguiente relación: $\theta = 4N_e\mu$ (θ = diversidad genética, N_e = tamaño efectivo poblacional, μ = tasa de mutación), o para el caso de un organismo haploide $\theta = 2N_e\mu$. Esta ecuación tiene sentido, pues dos individuos poseen mayor probabilidad de ser diferentes si la población es grande o la tasa de mutación es alta. Este tipo de predicciones hacen que la teoría neutra sea tan importante y central a la biología evolutiva moderna.

Adaptación a nivel molecular

El estudio de la adaptación molecular se ha realizado a partir de dos aproximaciones estadísticas diferentes. La primera es la distribución del polimorfismo en las secuencias de ADN y la segunda es la determinación de las sustituciones sinónimas y no-sinónimas en secuencias codificantes de ADN, utilizando el codón como mínima unidad evolutiva. Ambas aproximaciones hacen uso de las predicciones del modelo neutral para determinar la participación de la selección positiva.

Métodos basados en la distribución del polimorfismo

Los primeros trabajos desarrollados para el estudio de la selección a nivel molecular fueron los basados en la distribución del polimorfismo en la secuencia de ADN. Estos trabajos clásicos son la prueba de Tajima, el método de Fu-Li, que es una especie de modificación a la prueba de Tajima, la prueba de Hudson-Kreitman-Aguadè y el método de McDonald-Kreitman.

Para entender con mayor facilidad los fundamentos de las pruebas basadas en la distribución del polimorfismo en el ADN es necesario definir algunos parámetros básicos en genética de poblaciones que además son fundamentales en las predicciones de la teoría neutra.

Uno de estos parámetros es el de la diversidad genética, que puede ser descrita por dos estimadores: π (pi) y θ (teta). El primero (π) se refiere a la diversidad nucleotídica, que es el número de nucleótidos diferentes por sitio entre dos secuencias tomadas al azar (Nei y Li, 1979). El segundo (θ) es calculado a partir de la expresión $\theta = 4N_e\mu$ (Kimura, 1968). Sin embargo, es difícil tener la determinación exacta de los parámetros N_e (tamaño efectivo poblacional) y μ (tasa de mutación), por lo que una manera indirecta de estimar θ es utilizando el número total de sitios segregativos en un grupo de secuencias (un sitio segregativo es un sitio en donde las secuencias difieren), $\theta = K/a$, donde K es igual al número total de sitios segregativos en una muestra de secuencias dada y $a = 1 + 1/2 + \ldots + 1/n\text{-}1$ (n = número de secuencias en la muestra; Watterson, 1975; Tajima, 1983). Se deduce de estas expresiones que π se ve afectada mayormente por los alelos que poseen mayor frecuencia y es independiente del tamaño de la muestra, mientras que θ sí se ve afectada por el tamaño de la muestra y por la deriva génica, es decir, por los alelos poco frecuentes (Tajima, 1983, 1989).

La relación que existe entre ambos estimadores permite determinar si las secuencias se encuentran bajo el modelo neutral o se desvían del mismo. Si

ambos estimadores dan el mismo resultado en cuanto a variación genética, quiere decir que el polimorfismo observado es neutro y se encuentra distribuido aleatoriamente. En cambio, si existen diferencias entre ambos quiere decir que la selección está afectando alguno de ellos promoviendo su incremento o decremento: si existe selección positiva, ésta incrementará las frecuencias alélicas y eso se reflejará en el incremento de π; si existe un mayor número de alelos deletéreos en la muestra, θ se verá incrementada (Watterson, 1975; Tajima, 1983, 1989). De esta forma, si determinamos estadísticamente las diferencias entre ambos podremos detectar de manera indirecta la participación de la selección positiva o negativa en el mantenimiento del polimorfismo en las poblaciones.

Prueba de Tajima

Está basada en la detección de las diferencias entre los estimadores π y θ con la determinación de la *D* de Tajima, calculada a partir de la siguiente expresión:

$$D = \frac{\pi - \theta}{\sqrt{V(\pi - \theta)}}$$

Si *D* resulta negativa quiere decir que θ posee un valor mayor que π, lo que indica la presencia de mutaciones deletéreas. En cambio, si *D* resulta positiva quiere decir que π tiene un mayor valor que θ, indicación de que algunos alelos se encuentran bajo selección positiva (por ejemplo, selección balanceadora) incrementando sus frecuencias. Si *D* es igual a cero quiere decir que no existe diferencia alguna entre ambos estimadores y nos encontramos bajo equilibrio neutral (Tajima, 1983, 1989).

Método Fu-Li

Fue desarrollado con la misma lógica que la prueba de Tajima; la diferencia reside en que los autores consideran la distribución de las mutaciones en una genealogía de muestras tomadas al azar a partir de una población dada.

Suponen que las mutaciones antiguas tendrán que encontrarse con mayor probabilidad en las ramas más antiguas de la genealogía, mientras que las mutaciones originadas más recientemente en las ramas más nuevas. La parte más antigua de la genealogía consiste de las ramas internas y la parte más reciente de las ramas externas. Si existe selección negativa purificadora se observará un

exceso de las mutaciones en las ramas más externas. Por otra parte, si existe algún tipo de selección positiva se verá una disminución de las mutaciones en las ramas externas. Al comparar el número de mutaciones en ramas internas y externas con los valores esperados de acuerdo al modelo neutral se tendrá una prueba estadística sobre la participación de la selección(Fu y Li, 1993). La prueba consiste en aplicar la siguiente expresión:

$$G = \frac{n_e - n_i / a - 1}{\sqrt{V\,[n_e - n_i / (a - 1)]}}$$

donde
n_e = número de mutaciones en ramas externas
n_i = número de mutaciones en ramas internas

Ambos métodos parecen ser suficientemente efectivos para determinar si existe selección, aunque un estudio mostró que la prueba de Tajima es más efectiva (Simonsen *et al.*, 1995). Sin embargo ambos métodos asumen que las poblaciones presentan un equilibrio entre mutación y deriva génica por suficiente tiempo y toman como base el modelo de sitios infinitos, lo cual no siempre es válido sobre todo al estudiar secuencias codificantes de ADN donde las tres posiciones de los aminoácidos poseen distintas probabilidades de cambio.

Hudson-Kreitman-Aguadè (HKA)

Esta prueba es uno de los primeros intentos por determinar la participación de la selección a partir de las desviaciones del modelo neutral. Describe si existen diferencias entre el polimorfismo intrapoblacional y el polimorfismo entre diferentes poblaciones.

De acuerdo a los postulados de la teoría neutra, si todo el polimorfismo existente es neutro no deberían existir diferencias entre la variación genética dentro de los individuos de una especie de una población específica y entre poblaciones distintas de la especie, ya que todo el polimorfismo es originado al azar. En este caso se realiza una prueba donde se registra la correlación entre la diversidad genética intrapoblacional y entre poblaciones.

Se considera un número de loci (m) al azar en las poblaciones 1 y 2 y se toma para cada loci las secuencias n_1 y n_2 de las poblaciones 1 y 2, que presentan un número determinado de sitios segregativos o polimórficos K_{1i} del locus i

dentro de la población 1 y K_{2i} dentro de la población 2, y donde el número de diferencias entre las dos secuencias del locus *i* es D_i. Los parámetros K_{1i} y K_{2i} ($i = 1,\ldots, m$) son las medidas del polimorfismo intrapoblacional y la D_i es la medida de la variación entre las poblaciones (Hudson *et al.*, 1987).

Por otra parte, se realizan la determinación de los valores esperados de estos parámetros bajo el modelo neutro y el modelo HKA prueba si existen diferencias estadísticamente significativas entre los parámetros observados y los esperados (*E*) mediante la siguiente expresión:

$$X^2 = \sum_{i=1}^{m} [K_{li} - E(K_{li})] / V(K_{li}) + \sum_{i=1}^{m} [K_{2i})]^2 / V(K_{2i}) + \sum_{i=1}^{m} [D_t - E(D_t)] / V(D_t)$$

La determinación de los parámetros esperados y de las varianzas de estos parámetros se obtiene usando las siguientes expresiones a partir del modelo neutral:

$$E(K_{1i}) = a_1\theta_1$$
$$V(K_{1i}) = a_1\theta_1 + (b_1\theta_1)^2$$

donde

$$a_1 = 1 + 1/2 + \ldots + 1/n - 1$$
$$b_1 = 1 + 1/2^2 + \ldots + 1/(n-1)^2$$

Si la variación genética observada es muy diferente a la esperada por la teoría neutra, entonces se sugiere que la selección está participando en el mantenimiento del polimorfismo en las poblaciones. Los postulados del modelo HKA son los siguientes: 1) las generaciones son discretas; 2) todas las mutaciones son neutras; 3) el número de sitios en la secuencia es suficientemente grande para que cada mutación ocurra en un sitio que no ha sido mutado con anterioridad (modelo de sitios infinitos); 4) las mutaciones ocurren de manera independiente en cada locus; 5) el número de mutaciones por locus en cada generación tiene una distribución de Poisson; 6) no existe recombinación entre los loci; 7) los loci no se encuentran ligados genéticamente; 8) las especies 1 y 2 poseen los tamaños poblacionales *N* y *Nf*, respectivamente; 9) ambas especies son derivadas de un ancestro común en el tiempo *T* con un tamaño poblacional de *2N(1+f)/2* (promedio entre las poblaciones de las especies A y B; Hudson *et al.*, 1987). Estos postulados hacen que esta prueba no sea tan efectiva cuando se tienen muestras poblacionales pequeñas, cuan-

do las secuencias son muy divergentes entre sí y existe alta recombinación, y cuando se trabaja con muestras grandes de loci diferentes que se encuentran ligados genéticamente.

McDonald-Kreitman

Esta prueba está basada en principios muy sencillos, en los que se consideran las relaciones entre las diferencias sinónimas y no-sinónimas entre y dentro de las especies. La idea es parecida a la descrita por la prueba HKA, pero en este caso se prueba si el cociente de las sustituciones no-sinónimas a sinónimas entre las especies es el mismo que dentro de las poblaciones.

Por ejemplo, consideremos una región codificante de ADN y tomamos dos secuencias m_1 y m_2 a partir de dos especies cercanamente relacionadas 1 y 2. Sólo tomamos en cuenta los sitios que son variables entre las secuencias y descartamos los invariables. Dentro de estos sitios variables agrupamos dos categorías diferentes. Los sitios que son variables en un nucleótido (por ejemplo, G) en todos los miembros de la especie 1, y que difieren en otro nucleótido (por ejemplo, A) en todos los miembros de la especie 2, los denominamos sitios fijos. Son sitios fijos ya que son iguales para cada población en cada especie. Todos los demás sitios variables son considerados polimórficos. Estos sitios polimórficos serán diferentes dentro de las poblaciones de cada especie o entre las poblaciones de cada especie. Ambos tipos de sitios son clasificados en los que poseen sustituciones sinónimas y no-sinónimas. La hipótesis nula en la prueba de McDonald y Kreitman es la siguiente:

$$E\,(n_f)/Es_f = E\,(n_p)\,/\,Es_p$$

donde
n_f= sitios fijos con sustituciones no-sinónimas
s_f= sitios fijos con sustituciones sinónimas
n_p = sitios polimórficos con sustituciones no-sinónimas
s_p = sitios polimórficos con sustituciones sinónimas

Para ilustrar mejor las diferencias entre los parámetros estimados se puede construir una tabla de contingencia donde se comparen ambos tipos de sitios con sus respectivas sustituciones como lo realizado por McDonald y Kreitman (1991), al realizar las comparaciones entre tres especies de la mosca de la fruta *Drosophila* (tabla 1).

Tabla 1. Tabla de contingencia (2x2) para sitios fijos y polimórficos y sus respectivos tipos de sustitución, resultado de la comparación del gen de la Adh entre *D. melanogaster* (12 secuencias), *D. simulans* (6) y *D. yakuba* (24)

Sustituciones	Sitios fijos	Sitios polimórficos	Total
Sinónimas	sf (17)	sp (42)	sf + sp (59)
No-sinónimas	nf (7)	np (2)	nf + np (9)
Sumatoria	sf + nf (24)	sp + np (44)	68

Se aplicó una prueba de Fisher a estos resultados que mostró que existían diferencias estadísticamente significativas entre los cocientes de los sitios fijos y los sitios polimórficos (n_f/s_f = 7/17 y n_p/s_p = 2/42), siendo un poco mayor el cociente de los sitios fijos que el de los polimórficos, lo que sugiere que la selección positiva promovía un pequeño incremento en las sustituciones no-sinónimas entre especies.

Una debilidad de este método es que no contempla la existencia de sustituciones múltiples en un mismo sitio; algunos estudios posteriores sugieren que si esto se toma en cuenta no se puede rechazar la neutralidad para el gen de *Adh* en las tres especies de la mosca de la fruta (Whittam y Nei, 1991). Además, ya que la teoría neutra incluye la participación de la selección negativa y mutación, la prueba de McDonald y Kreitman no necesariamente descarta la neutralidad. Por lo mismo es necesario ser cuidadoso en la interpretación de los resultados obtenidos utilizando este método.

Métodos basados en sustituciones nucleotídicas

En cada codón, que corresponde a un tipo de aminoácido, se encuentran tres posiciones denominadas primera, segunda y tercera posición, correspondientes a las tres bases que lo integran. Cada una de estas posiciones se ve afectada por distintas tasas de mutación, por lo que ni la mutación ni la selección es uniforme en cada uno de los codones.

Por ejemplo, de acuerdo al código genético si un cambio es introducido en la primera posición en su mayoría será un cambio sinónimo y un porcentaje bajo corresponderá a un cambio no sinónimo, es decir se mantendrá codificando para el mismo tipo de aminoácido. Si un cambio se introduce en la

tercera posición, en su mayoría corresponderá a un cambio sinónimo. Por otro lado, si un cambio se introduce en la segunda posición, éste repercutirá directamente en el tipo de aminoácido para el cual está codificando. Por lo tanto, la segunda es la única posición que nos brinda la posibilidad de evaluar un cambio que potencialmente podría ser adaptativo y que estará sujeto directamente a la selección natural.

Existen numerosos métodos para la estimación de las tasas de sustitución y pueden ser clasificados en tres grupos: a) métodos evolutivos, b) métodos basados en el modelo de Kimura-2 parámetros, y c) métodos de máxima verosimilitud basados en un modelo de sustitución de codones.

Métodos evolutivos

1) Detectando selección positiva con el método de Nei-Gojobori
Dentro de los métodos evolutivos, éste y su modificación (ver más adelante) son los más utilizados. En el método de original las sustituciones sinónimas y no-sinónimas se determinan calculando el número observado de sustituciones sinónimas y no-sinónimas y el número de sitios con sustituciones potencialmente sinónimas y no-sinónimas. Los sitios potencialmente sinónimos y no-sinónimos son descritos a partir de cada codón asumiendo que existe igual probabilidad de cambio para todos los nucleótidos. El número de sitios potencialmente sinónimos (*s*) y no-sinónimos (*n*) para un codón específico es calculado con la siguiente relación:

$$S = \Sigma_{i=1}^{3} f_i$$

$$n = 3 - s$$

donde f_i es la proporción de cambios sinónimos (cociente del número de cambios sinónimos y no-sinónimos excluyendo mutaciones sin sentido en la posición *i*). Por ejemplo, en el caso del codón de la fenilalanina:

T T T
s = 0 + 0 +1/3

porque todos los cambios en la primera y segunda posición generan un cambio no-sinónimo y en la tercera posición 1 de cada 3 posibles cambios produce una sustitución sinónima (TTC; tomado de Nei y Kumar, 2000).

Como todos los cambios restantes originan una sustitución no-sinónima,

$n = 3 - 1/3 = 8/4$

En el caso de que cualquier cambio origine un codón de término, éste no será tomado en cuenta para el cálculo.

Para obtener el número total de sitios sinónimos (*S*) y no-sinónimos (*N*) en una secuencia entera usamos la fórmula:

$S = \sum_{j=1}^{c} s_j$

donde s_j= valor de *s* para el codón *j*
C= número total de codones
$S + N = 3C$ (número total de nucleótidos comparados)
Aplicando esta fórmula en un ejemplo, tenemos la comparación de dos codones TTG- Leu y AGA- Arg, cada uno en una secuencia dada. Para la transición de Leu a Arg existen 6 escenarios evolutivos posibles:

1) TTG (Leu) - ATG (Met) - AGG (Arg)- AGA (Arg)
2) TTG (Leu) - ATG (Met) - ATA (Ile) - AGA (Arg)
3) TTG (Leu) - TGG (Trp) - AGG (Arg) - AGA (Arg)
4) TTG (Leu) - TGG (Trp) - TGA (Alto) - AGA (Arg)
5) TTG (Leu) - TTA (Leu) - ATA (Ile) - AGA (Arg)
6) TTG (Leu) - TTA (Leu) - TGA (Alto) - AGA (Arg)

Como los escenarios 4 y 6 involucran un codón de término, no son tomados en cuenta para el estudio. El número de sustituciones sinónimas en los escenarios 1, 2, 3 y 5 son 1, 0, 1 y 1, respectivamente, mientras que para el caso de las sustituciones no-sinónimas es 2, 3, 2, y 2, respectivamente. Ya que asumimos que los 4 escenarios son igualmente probables, tenemos que S*d*= 3/4 y *Nd*= 9/4 (tomado de Kumar y Nei, 2000). Podemos obtener las sustituciones sinónimas y no-sinónimas en la comparación de dos secuencias sumando los valores de todos los codones. Esto es,

$Sd = \sum_{j=1}^{c} sd_j$

$$Nd = \sum_{j=1}^{c} ndj$$

donde sd_j y nd_j corresponden a los números de las diferencias sinónimas y no-sinónimas para el codón j y C es el número de codones comparados.

Por lo tanto, $Sd + Nd$ es igual al número total de diferencias entre dos secuencias comparadas. A partir de esto podemos estimar la proporción de cambios sinónimos (*PS*) y no-sinónimos (*PN*) como:

$$PS = \frac{Sd}{S}$$

$$PN = \frac{Nd}{N}$$

donde *S* y *N* se refieren al número promedio de sitios sinónimos y no-sinónimos para dos secuencias comparadas.

A partir del modelo de Jukes-Cantor (1969) para calcular el número de diferencias de nucleótidos (*d*) entre dos secuencias, tenemos la siguiente relación:

$$d = - (3/4)\ 1n\ [(1- (4/3)p]$$

donde

$p = 1 - q$ es la proporción de nucleótidos diferentes entre dos secuencias y *q* es el número de nucleótidos idénticos entre dos secuencias.

Podemos usar esta misma ecuación para determinar el número de sustituciones sinónimas y no-sinónimas por sitio remplazando *p* por *PS* o *PN*. Esta ecuación nos da buenos resultados, siempre y cuando las frecuencias de las bases A, T, G y C sean casi iguales y no exista una desviación en el cociente de transiciones y transversiones, es decir, que ambos cambios sucedan con la misma probabilidad (Ota y Nei, 1994). Para el cálculo de la varianza de *dS* y *dN* podemos usar la misma fórmula de la varianza basada en el método de Jukes-Cantor (Ota y Nei, 1994):

$$V(ds) = \frac{V(PS)}{(1 - 4/3\ PS)}$$

$$V(dN) = \frac{V(PN)}{(1 - 4/3\ PN)^2}$$

donde,

$$V(PS) = \sum_{j=1}^{c}(sdi - PSsi)^2 / S^2$$

$$V(PN) = \sum_{j=1}^{c}(ndi - PNni)^2 / N^2$$

Otra forma de calcular sus varianzas es el método de *bootstrap*, siempre y cuando las muestras de *Sd*, *Sn*, *S* y *N* sean suficientemente grandes (Nei y Kumar, 2000).

Como vimos con anterioridad, para detectar selección positiva a nivel molecular es necesario mostrar que *dN* es significativamente mayor que *dS*. Un método muy simple es ver las diferencias entre las tasas de sustitución sinónima y no sinónima $D = dN - dS$ y su varianza $V(D)$ (Nei y Kumar, 2000). Siempre y cuando nuestra muestra sea grande la distribución aproximada de *D* es la normal y, en este caso en particular, la $V(D) = V(dN) + V(dS)$, por lo que tendríamos la siguiente relación a usar en una prueba estadística de *Z*:

$$Z = D / s(D)$$

donde

$$s(d) = V(D)^{1/2}$$

Como estamos interesados en saber si $dN > dS$ y se analiza como una prueba de una cola con un infinito número de grados de libertad. La varianza de *D* también puede ser calculada a mediante una prueba de *bootstrap*, en donde se toma como unidad de muestreo sucesivo los codones en lugar de los nucleótidos de la secuencia y se procede a realizar múltiples repeticiones

de los parámetros *Psb*, *Pnb*, *dsb*, *dnb* de *PS, PN, dS, dN* y se obtienen las varianzas de estos valores.

También podemos calcular el error estándar de *D* mediante *bootstrap* y luego usar una prueba de *Z* (si la muestra es grande) (Efron y Tibshirani, 1993). Una forma más adecuada en el caso de que la secuencia de nucleótidos sea pequeña es calcular directamente *dS* y *dN* y sus varianzas y usar la hipótesis nula de que ambas son iguales ($dN = dS$) con una prueba de *Z*. Otra forma es construir una tabla de contingencia para las sustituciones sinónimas y no-sinónimas calculadas directamente de la secuencia y realizar una prueba de Fisher (Zhang *et al.*, 1997). Si el valor de *P* es menor a 0.05 entonces la hipótesis nula de evolución neutral es rechazada.

Método modificado de Nei-Gojobori

Al realizar el cálculo del número de sitios sinónimos y no-sinónimos el método de Nei-Gojobori asume que la sustitución de los 4 nucleótidos sucede de forma azarosa, es decir con la misma probabilidad para cada uno. En la práctica esto no es necesariamente cierto ya que puede existir un mayor número de sustituciones de un tipo de nucleótido u otro. El cociente de transiciones (cambio de purina por purina o pirimidina por pirimidina) es usualmente mayor que el de transversiones (cambio de purina por pirimidina y viceversa).

En este caso, se espera que el número de sitios potenciales que producen sustituciones sinónimas sea mayor que el determinado por el método de Nei-Gojobori, ya que las transiciones en la tercera posición son en su gran mayoría sinónimas y serán más frecuentes. Para corregir este detalle se propuso modificarlo utilizando como base el modelo de dos parámetros de Kimura (Kimura-2; Kimura, 1980; Ina, 1995), en este modelo la proporción de transiciones del total de cambios es dada por la siguiente relación:

$$\alpha / \alpha + 2\beta = R/1 + R$$

α = transiciones, cada nucleótido puede tener un cambio transicional, por ejemplo A (purina) a G (purina).
β = transversiones, 2β tomando en cuenta que cada nucleótido puede tener dos cambios transversionales, por ejemplo, A (purina) puede cambiar a T (pirimidina) o C (pirimidina).

R = cociente de transición/transversión, que es igual a 0.5 cuando las transiciones y las tranversiones se producen con la misma probabilidad.

Teóricamente, la modificación al método de Nei-Gojobori se apega más al proceso real de sustituciones en una secuencia de ADN. Sin embargo, determinar el cociente de transiciones/transversiones (*R*) en los sitios sinónimos o no-sinónimos es complicado y usualmente se cae en sobreestimaciones de cambios sinónimos. Por otra parte, cuando se sobreestima este cociente se puede concluir que las sustituciones no-sinónimas son significativamente mayores a las sinónimas, cuando no es el caso.

Hay que señalar que la mayor parte de las veces los patrones de sustitución son más complicados que los descritos por el modelo de Kimura-2, y bajo ciertas condiciones tanto el método de Nei-Gojobori como su modificación pueden sobreestimar las sustituciones sinónimas sobre las no-sinónimas, haciendo difícil la detección correcta de selección positiva. Lo recomendable es usar ambos métodos, y si ambas versiones detectan la presencia de selección positiva se tiene una mayor seguridad acerca de las conclusiones derivadas de este análisis.

Métodos basados en el modelo de dos parámetros de Kimura (Kimura, 1980)

Además de la modificación al modelo de Nei-Gojobori, existen otros métodos que intentan detectar la selección basados en el modelo de Kimura-2. Entre los más importantes encontramos el método de Li-Wu-Luo, el de Pamilo-Bianchi-Li y el de Cameron y Kumar.

Li-Wu-Luo

Li y sus colaboradores (1985) desarrollaron otro método considerando la degeneración del código genético. Los sitios en los codones pueden ser clasificados en 4-veces degenerados (L_4), 2-veces degenerados (L_2) y 0-veces degenerados (L_0). Un sitio es reconocido como 4-veces degenerado si todos los cambios posibles en ese sitio son sinónimos, 2-veces degenerado si uno de tres cambios o dos de tres cambios es sinónimo y 0-veces degenerado si todos los cambios son no-sinónimos o se refieren a mutaciones sin sentido que generan codones de término.

Por ejemplo, la mayoría de los cambios en la tercera posición son 4-veces degenerados, mientras que todos los cambios en la segunda posición son 0-veces degenerados. Bajo esta regla podemos calcular los tres tipos de cambios L_0, L_2, y L_4 para dos secuencias que quieran ser comparadas. Se comparan las dos secuencias codón por codón y se clasifica cada diferente nucleótido según sea una transición o transversión. Entonces se calcula la proporción de transiciones (*Pi*) y transversiones (*Qi*) en cada clase (*i*) de nucleótido (*i*= 0, 2 ó 4) tomando en cuenta la probabilidad de verosimilitud de ocurrencia de cada aminoácido (se denomina en inglés *likelihood occurrence*, es un parámetro de probabilidad de que se encuentre presente un determinado aminoácido).

Ellos estiman la proporción de cambios transicionales y de transversiones con la siguiente relación:

$$Ai = 1/2\ 1n\ (a_i) - 1/4\ (b_i)$$

$$B_i = 1/2\ 1n\ (b_i)$$

donde

$$a_i = 1/(1 - 2Pi - Q_i)$$

$$b_i = 1/(1\text{-}2Q_i)$$

Pi = número de transiciones tomando en cuenta la clase de sitio (0, 2 ó 4)
Qi = número de transversiones tomando en cuenta la clase de sitio (0, 2 ó 4).

Sabemos que la mayoría de las sustituciones en los sitios 4-veces degenerados son sinónimas y todas las sustituciones en los sitios 0-veces degenerados son no-sinónimas. Sin embargo, para los sitios 2-veces degenerados las transiciones (A_2) son en su mayoría sinónimas, mientras que las transversiones (B_2) son en su mayoría no-sinónimas. Si suponemos que la probabilidad de sustitución de cada uno de los 4 nucleótidos (A, G, C y T) se lleva a cabo con la misma frecuencia, los autores de este método sugirieron que 1/3 de los cambios en los sitios 2-veces degenerados son potencialmente sinónimos y 2/3 son potencialmente no-sinónimos. Bajo estas ideas propusieron que *dS* y *dN* pueden ser estimados con la siguiente relación:

$$\mathrm{dS} = \frac{3\ [L_2 A_2 + L_4\ (A_4 + B_4)\ ]}{L_2 + 3L_4}$$

$$dN = \frac{3\,[L_0(A_0 + B_0 + L_2 B_2\,]}{3L_0 + 2L_2}$$

A su vez, estas fórmulas se basan en ciertos supuestos, el primero y más evidente es que el tipo de sitio en una determinada secuencia puede no corresponder al mismo sitio o sitio homólogo en la secuencia contra la cual se está comparando. Es decir, en una secuencia puede tratarse de un sitio 4-veces degenerado, mientras que en su equivalente este sitio puede ser un sitio 2-veces degenerado, lo que sucede con frecuencia si trabajamos con secuencias muy divergentes. En este caso en particular la mitad de los sitios son considerados 2-veces degenerados y la otra mitad 4-veces degenerados.

El segundo supuesto es que los cambios sin sentido son considerados cambios no-sinónimos (lo que es totalmente incorrecto). Se sabe que este tipo de cambios suceden con una probabilidad de 4%, por lo tanto, se espera que el método de Li-Wu-Luo sobreestime las sustituciones no-sinónimas (*dN*). Aunado a esto tenemos casos específicos de ciertos aminoácidos donde surgen problemas en las determinaciones, tal es el caso de la leucina donde en la tercera posición de sus tres codones (2-veces degenerados), algunas transversiones son sinónimas.

Otro caso es el de la arginina, en donde las transiciones en las primeras posiciones de 4 de sus codones (2-veces degenerados) son no-sinónimas con excepción de un codón en donde el cambio genera una mutación sin sentido.

A pesar de estos detalles este método proporciona estimados parecidos al de Nei-Gojobori con precisión, siempre y cuando se trate del análisis de muchos codones (>100) y secuencias poco divergentes entre sí.

Pamilo-Bianchi-Li

Este método es una extensión del anterior (Pamilo y Bianchi, 1993; Li, 1993). Los autores notaron que las transiciones que resultan en cambios sinónimos suceden sólo en los sitios 2-veces y 4-veces degenerados, como las transversiones en los sitios 4-veces degenerados son también sinónimas, el número total de sustituciones sinónimas por sitios sinónimos se estima de la siguiente manera:

$$dS = (L_2 A_2 + L_4 A_4)(L_2 + L_4) + B_4)$$

Y por lo tanto las sustituciones no-sinónimas,

$$dN = A_0 + (L_0 B_0 + L_2 B_2) / (L_0 + L_2)$$

Cameron y Kumar

Ninguno de los métodos anteriores toma en cuenta casos específicos como la arginina, aunque en primera instancia esto no parece tener importancia, pues se trata de un solo aminoácido, pero en algunos genes de mamíferos (como la protamina) es un problema real, ya que sus proteínas son precisamente ricas en arginina.

Utilizando las mismas fórmulas del método de Pamilo-Bianchi-Li, Cameron (1995) y Kumar *et al.* (2000) dividieron los sitios 2-veces degenerados en sitios 2-veces degenerados simples y complejos. Los sitios simples se refieren a sitios en que las transiciones dan lugar a un cambio sinónimo y dos transversiones generan sustituciones no-sinónimas o sin sentido. Todos los casos restantes se incluyen en los sitios 2-veces degenerados complejos.

Métodos de verosimilitud con modelos de sustitución de nucleótidos

Como alternativa a los métodos descritos anteriormente surgen los modelos basados en la aproximación de máxima verosimilitud. En estos modelos se pretende incluir algunos detalles que los demás no incluían como es el caso de que una clase de sitio (0-veces degenerado, por ejemplo) haya cambiado en el tiempo, además de la inclusión de múltiples tasas de sustitución para un sitio determinado (Muse y Gaut, 1994; Goldman y Yang, 1994).

Dos modelos surgieron al mismo tiempo utilizando esta aproximación. Uno de ellos es el trabajo llevado a cabo por Muse y Gaut (1994), utilizando como modelo el genoma del cloroplasto, y el segundo fue desarrollado por Goldman y Yang (1994). Ambos se basan en un modelo específico de sustitución de codones y en la determinación de la mayoría de los parámetros del modelo por el método de Máxima Verosimilitud descrito por Felsenstein (1981). En estos trabajos se supone que cada sitio en la secuencia evoluciona de manera independiente de sus sitios vecinos y de los sitios homólogos en las otras secuencias. De esta forma el parámetro de máxima verosimilitud de

cada sitio (algo así como la probabilidad de observar dicho nucleótido en un lugar específico de la secuencia) es el producto de la máxima verosimilitud de cada uno de los sitios.

Comenzaremos por detallar el funcionamiento de estos modelos analizando los trabajos desarrollados por Goldman y Yang (1994), Yang y Nielsen (2000) y Nielsen y Yang (1998), entre algunos, ya que son los más desarrollados y sobre los cuales se basan en la actualidad la mayoría de los trabajos de estudio de adaptación molecular con el método de máxima verosimilitud. Los autores definen el parámetro *W* como el cociente de sustituciones no-sinónimas y sinónimas (*dN/dS*), que al igual que en los demás modelos mide la relación entre los dos tipos de sustituciones. Si un cambio en un aminoácido es neutral se fijará con la misma probabilidad que una mutación sinónima y $W = 1$. Si el cambio en el aminoácido es deletéreo la selección negativa o purificadora reducirá su probabilidad de fijación y $W < 1$. Sólo cuando el cambio de un aminoácido tiene una ventaja selectiva se fija con una mayor probabilidad que una mutación sinónima y $W > 1$. Por lo tanto, un valor de *W* significativamente mayor a uno es una evidencia de selección positiva. El modelo de sustitución de codones con el cual determinan la tasa de sustitución del codón *i* al *j* $(i \neq j)$ está dado por las siguientes condiciones:

$$qij \begin{cases} 0, \text{ si } i \text{ y } j \text{ difieren en una posición} \\ \pi j, \text{ para transversiones sinónimas} \\ \kappa\pi j, \text{ para transiciones sinónimas} \\ \omega\pi j, \text{ para transversiones no-sinónimas} \\ \omega\kappa\pi j, \text{ para transiciones no-sinónimas} \end{cases}$$

donde
πj – frecuencia del codon *j* en equilibrio (tomado de la frecuencia del codón en la muestra en estudio)
κ - cociente de transiciones y transversiones determinado a partir de la muestra
ω - *dN/dS*

Si se quiere considerar la sustitución a través de un tiempo *t* la probabilidad de transición está dada por:

$$P(t) = \{pij\ (t)\} = e^{Qt}$$

donde

$p_{ij}(t)$es la probabilidad de que el codón *i* se convierta en el *j* en un tiempo *tQ* es igual a p_{ij}.

Después de que se realizan las determinaciones de los parámetros correspondientes se procede a realizar pruebas estadísticas que nos permitirán determinar si la tasa de sustitución no sinónima es significativamente mayor que la sinónima.

En el caso de los métodos basados en máxima verosimilitud se utiliza una prueba de máxima verosimilitud con modelos alternativos donde se fijan varios valores a *W* y se comparan con los obtenidos en la muestra. Dos veces el logaritmo de la función de verosimilitud obtenida para cada modelo se compara con una distribución de χ^2 con un grado de libertad y se prueba la hipótesis nula de que *W* es igual o mayor a uno (Goldman y Yang, 1994; Nielsen y Yang, 1998; Yang y Nielsen, 2000; Yang y Bielawski, 2000). Si éste es el caso, entonces se puede decir con una alta probabilidad que existe selección positiva actuando en el gen en estudio.

Como podemos ver, hemos trabajado definiendo el cociente de *dN/dS* para toda la secuencia del gen y lo que obtenemos siempre es un valor promedio de todos los sitios de esta secuencia. En este caso sólo detectamos la selección si ésta es en promedio mayor a uno, lo cual no sucede en la mayoría de los casos (Goldman y Yang, 1994; Nielsen y Yang, 1998; Yang y Nielsen, 2000; Yang y Bielawski, 2000). Sin embargo, esto no quiere decir que no existan sitios particulares dentro de la secuencia que se encuentren seleccionados positivamente. Si lo que pretendemos es detectar cuáles aminoácidos dentro de la secuencia se encuentran bajo selección positiva con el modelo de máxima verosimilitud, entonces la prueba de selección involucra dos pasos.

El primero es probar si existen sitios dentro de la secuencia donde *W* sea mayor que 1, lo que se lleva a cabo con la determinación del cociente de verosimilitud comparando un modelo en donde se tengan sitios que posean una $W > 1$, con otro modelo donde no existan esos sitios; el segundo paso implica la aplicación del teorema de Bayes para identificar los sitios donde realmente existe selección positiva (Goldman y Yang, 1994; Nielsen y Yang, 1998; Yang y Nielsen, 2000; Yang y Bielawski, 2000).

Para poder tener una mayor comprensión del método, haremos un paréntesis para explicar con un modelo cómo es que funcionan estas pruebas (tomado de Yang y Bielawski, 2000). Supongamos que tenemos una población hipotética dividida en dos grupos G1 y G2 que corresponden al 60% y 40% de la población total, respectivamente. En el grupo G1 ocurre un tipo de en-

fermedad en una proporción de 1% y en el grupo G2 de 0.01%. Supongamos ahora que hacemos un muestreo al azar de 100 individuos de la población total (incluyendo G1 y G2), ¿cuál es la probabilidad de que 3 de ellos estén enfermos?

La probabilidad (D) de que un individuo tomado al azar de la población esté enfermo es un promedio entre los dos grupos G1 y G2, y está definida por:

p = P(D) = P(G1) x P(D G1) + P(G2) x P(D G2) = 0.6 x 0.01 + 0.4 x 0.001 = 0.0064

De igual manera la probabilidad de que el individuo no esté enfermo es:

$P(D_2)$= P(G1) x $P(D_2$ G1) + P(G2) x $P(D_2$ G2) = 0.6 x 0.99 + 0.4 x 0.999 = 0.9936

lo que es lo mismo que $1 - p$ (de la primera ecuación).

La probabilidad de que 3 de cada 100 individuos posean la enfermedad está dada por la probabilidad:

$$P = \frac{100!\, p^3\, (1 - p)^{97}}{3!\, x\, 97!} = 0.0227$$

Si esta última ecuación involucra un parámetro desconocido, por ejemplo, la probabilidad de aparición de la enfermedad en el grupo G1, éste puede ser estimado maximizando esta fórmula. En este caso, la ecuación nos da la probabilidad de observar el dato de la muestra y se le denomina la función de verosimilitud.

La segunda cuestión involucra realizar el razonamiento de forma inversa, es decir, calcular la probabilidad de que un individuo de la muestra tomada al azar que se encuentre enfermo pertenezca al grupo G1. Es aquí que usamos el teorema de Bayes:

$$P(\text{G1 D}) = P(\text{G1}) \text{ x } P(\text{D G1})/P(\text{D}) = 0.6 \text{ x } 0.01/0.0064 = 0.94$$

Por lo tanto, la probabilidad de que un individuo enfermo pertenezca al grupo I es mayor, de la misma forma la probabilidad de que un individuo sano pertenezca a este grupo es mayor.

Para el caso del estudio de la selección positiva, la *D* del ejemplo corresponde a un sitio de la secuencia y *Gi* es la clase de sitio *i* a la que pertenece con un determinado valor de *Wi*. La probabilidad de observar un dato determinado en un sitio es un promedio de las clases de sitios. Entonces, el producto de las probabilidades de clases de esos sitios constituye la verosimilitud a partir de las cual se estiman los parámetros desconocidos, tal como la distribución de *W*

sobre los sitios y la longitud de las ramas de los árboles filogenéticos o genealogías. Después de que éstos parámetros son estimados, se utiliza el teorema de Bayes para el cálculo de la probabilidad de que cualquier sitio pertenece a una clase determinada de valores de W, y dentro de ellos se encuentre el modelo de selección positiva cuyos valores serán mayores a uno.

Una de las ventajas de los modelos basados en máxima verosimilitud es que no necesitan de las reconstrucciones filogenéticas, además de que pueden incluir ciertos parámetros característicos del proceso evolutivo del ADN, tales como el cociente de transiciones/transversiones y el uso de codones.

Métodos basados en reconstrucciones filogenéticas

Dentro de los métodos que intentan determinar la selección a nivel intragénico llaman la atención los trabajos que incluyen una reconstrucción de los posibles ancestros basándose en una filogenia o genealogía determinada. Uno de los trabajos más sistemáticos es realizado por Suzuki y Gojobori (1999). Los autores calculan el número de sitios sinónimos y no-sinónimos a lo largo de un árbol filogenético construido a partir del modelo de distancia del vecino más cercano (*neighbor-joining*), y luego prueban si la proporción de sustituciones no-sinónimas difiere del modelo neutral (*i.e.* $W = 1$). Las tasas de sustitución las calculan a partir del modelo de Nei-Gojobori. Para cada sitio en el codón se infiere su secuencia ancestral a partir de cada nodo del árbol.

Otro trabajo basado en reconstrucciones genealógicas es el realizado por Fitch y colaboradores (1997) donde básicamente usan el modelo de máxima parsimonia para la reconstrucción de una genealogía y a partir de ella calculan los cambios en cada sitio de cada codón a lo largo de las ramas de las genealogías obtenidas asumiendo que las tasas de sustitución son constantes en todos los sitios de los codones (lo que en general no es cierto). Estos métodos necesitan de muestras grandes para ser eficientes y no tener varianzas muy grandes, recordando que a partir de las reconstrucciones filogenéticas ya existen errores dependiendo del modelo evolutivo usado, así es que hay que tener cuidado en su implementación.

En conclusión, la mayoría de los métodos presentados aquí son vigentes en la literatura de investigación de los procesos evolutivos moleculares. Todos son efectivos dependiendo del tipo de datos y las características de la muestra a usar, sobre todo si trabajamos con secuencias muy conservadas y poco divergentes entre sí a través del tiempo. Sin embargo, lo que sí es claro es que cuando trabajamos con secuencias poco conservadas y divergentes

entre sí, los métodos de máxima verosimilitud y las reconstrucciones basadas en modelos de distancia resultan ser más efectivos que los demás (Zhang y Nei, 1997). Por ejemplo, la probabilidad de reconstrucción correcta de las secuencias ancestrales, utilizando como modelo la secuencia de la lisozima C, es de 98.7% para el método de máxima verosimilitud contra 91.3% para los métodos de máxima parsimonia (Yang *et al.*, 1995).

En la tabla 2 se comparan las tasas de sustitución obtenidas con varios de los métodos anteriormente descritos en la subunidad α_2 de las globinas entre humanos y orangutanes (142 codones; tomado de Yang y Bielawski, 2000) y el gen mitocondrial para la NADH deshidrogenasa de humanos y chimpancés (603 codones; modificado de Nei y Kumar, 2000).

Como podemos observar no existe una diferencia importante entre los distintos modelos para estos genes.

Genes donde se ha detectado selección positiva

Hemos intentado dar un panorama de los diferentes métodos utilizados en la actualidad para el estudio de la selección positiva a nivel molecular, aunque no existe un consenso acerca de cuál es el mejor de ellos. En general existe una amplia gama de trabajos y autores que usan indistintamente todos los métodos descritos en el presente trabajo con resultados muy parecidos.

Lo más importante es conocer la muestra de datos que vamos a utilizar y todas las peculiaridades que exhibe, como el cociente de transversiones/transiciones y el uso de codones. También es importante tener una reconstrucción filogenética adecuada. De esta forma aseguraremos que el modelo que escojamos para determinar las tasas de sustitución y calcular la relación *dS*/*dN* es el que más se ajusta a nuestros datos. Otra sugerencia es la comparación de resultados utilizando varios métodos y hacer acopio de toda la información de evolución molecular y biológica a la que tengamos acceso, lo que siempre garantizará la obtención de mejores resultados.

En la tabla 3 se resume la mayoría de los ejemplos registrados en los que se ha detectado selección positiva mediante la determinación del cociente *dN*/*dS* (modificado de Yang y Bielawski, 2000)

En la literatura la mayoría de los trabajos señalan que la mayor parte de la evolución molecular está dominada por procesos neutros y selección negativa o purificadora, como Kimura (1983) y otros autores (Ohta, 1992; Li, 1997) han discutido. Pero ¿qué sucede cuando estudiamos los genomas completos? ¿acaso la selección actúa de la misma forma que a nivel génico e intragénico?

Tabla 2. Tasas de sustitución de nucleótidos de la subunidad α_2 de las globinas y de la NADH deshidrogenasa determinadas con distintos métodos

Método	dN	dS	dN/dS	Referencia
globinas α2				
Nei-Gojobori	0.0095	0.0569	0.168	Nei y Gojobori, 1986
Li	0.0104	0.0517	0.201	Li, 1993
Ina	0.0101	0.0523	0.193	Ina,1995
Yang y Nielsen	0.0083	0.1065	0.078	Yang y Nielsen, 2000
NADH deshidrogensasa				
Nei-Gojobori	0.0379	0.4151	---	Kumar y Nei, 2000
Li-Wu-Luo	0.0378	0.4277	---	Kumar y Nei, 2000
Modificación a Nei-Gojobori	0.0438	0.2730	---	Kumar y Nei, 2000
Ina	0.0438	0.3031	---	Kumar y Nei, 2000
Pamilo-Bianchi-Li	0.0438	0.3018	---	Kumar y Nei, 2000
Cameron-Kumar	0.0438	0.3018	---	Kumar y Nei, 2000
Goldman-Yang	0.0442	0.2872	---	Kumar y Nei, 2000

Tabla 3. Genes involucrados en actividades celulares de varias clases que se encuentran bajo selección positiva según diferentes métodos moleculares

Gen y actividad celular	Organismo	Método	Referencia
Sistema inmune			
Quitinasas clase I	*Arabis* y *Arabidopsis*	Li-Wu-Luo	Bishop *et al.*, 2000.
Colicinas	*Escherichia coli*	Nei-Gojobori	Riley, 1993.
Defensinas	Roedores	Nei-Gojobori	Hughes y Yeager, 1997.
Inmunoglobulinas VH	Mamíferos	Nei-Gojobori	Tanaka y Nei, 1989.
Genes del Complemento de Máxima Histocompatibilidad	Mamíferos	Nei-Gojobori	Hughes y Nei, 1988.

(Continúa)

Tabla 3. Genes involucrados en actividades celulares de varias clases que se encuentran bajo selección positiva según diferentes métodos moleculares *(continúa)*

Gen y actividad celular	Organismo	Método	Referencia
Inhibidor de la poligalacturonasa y legumbres	Dicotiledóneas	---	Stotz *et al.*, 2000.
Genes de grupo sanguíneo RH (RH50)	Primates y roedores	Nei-Gojobori	Kitano *et al.,* 1998.
Ribonucleasas	Primates	Nei-Gojobori	Zhang *et al.,* 1998.
Gen de la transferina	Salmones	Nei-Gojobori	Ford *et al.,* 1999.
Interferon- ω tipo I	Mamíferos	Nei-Gojobori	Hughes, 1995.
Inhibidor de la proteinasa ω-1	Roedores	Nei-Gojobori	Goodwin *et al.,* 1996.
Evasión de la respuesta a sistema inmune			
Gen de la cápside	Virus FMDV	ML*	Haydon *et al.,* 2001.
Genes CSP, TRAP, MSA-2 y PF83	*Plasmodium falciparum*	Nei-Gojobori	Hughes y Hughes, 1995.
Región codificadora del antígeno delta	Virus de la hepatitis D	Cameron	Wu et al., 1999.
Gen de la envoltura de la cápside	Virus HIV	Ina	Yamaguchi-Kabata y Gojobori, 2000.
Glicoproteína gH	Virus de la pseudorabia	Nei-Gojobori	Endo *et al.,* 1996.
Hemaglutinina	Virus humano de la influenza tipo A	---	Fitch *et al.,* 1997.
Genes del antígeno de la invasión del plásmido	*Shigella*	Nei-Gojobori	Endo *et al.,* 1996.
Antígeno-1 de la superficie del merozoito (MSA-1)	*Plasmodium falciparum*	Nei-Gojobori	Hughes, 1992.
Proteína de la membrana externa	Chlamydia	Nei-Gojobori	Endo *et al.,* 1996.
Genes de la poligalacturonasa	Hongos patógenos	---	Stotz *et al.,* 2000.
Porina 1 (porB)	*Neisseria gonhorreae* y *N. meningitidis*	Nei-Gojobori	Smith *et al.,* 1995.
Glicoproteínas S y HE	Coronavirus del ratón	Nei-Gojobori	Baric *et al.,* 1997.
Sigma 1	Reovirus	Nei-Gojobori	Endo *et al.,* 1996.
Gen determinante de la virulencia	Yersinia	Nei-Gojobori	Endo *et al.,* 1996.

(Continúa)

Tabla 3. Genes involucrados en actividades celulares de varias clases que se encuentran bajo selección positiva según diferentes métodos moleculares *(continúa)*

Gen y actividad celular	Organismo	Método	Referencia
Reproducción			
Proteína de 18 kDa involucrada en la fertilización	*Haliotis* (Abalone)	Nei-Gojobori	Vacquier *et al.,* 1997.
Gen Acp26Aa	*Drosophila*	Li-Wu-Luo	Tsaur y Wu, 1997.
Proteína que une al andrógeno	Roedores	Li	Karn y Nachman, 1999.
Hormona de puesta de huevos	*Aplysia californica*	Nei-Gojobori	Endo *et al.,* 1996.
Gen de caja homeótica Ods	Drosophila	---	Ting *et al.,* 1998.
Gen de caja homeótica Pem	Roedores	Li	Sutton y Wilkinson, 1997.
Protamina P1	Primates	Nei-Gojobori	Rooney y Zhang, 1999.
Lisina del esperma	*Haliotis* (Abalone)	Nei-Gojobori	Vacquier *et al.,* 1997.
RNAsa S	Rosaceae	Nei-Gojobori	Ishimizu *et al.,* 1998.
Gen de la diferenciación testicular Sry	Primates	Pamilo-Bianchi-Li	Pamilo y O'Neill, 1997
Digestión			
Caseína *K*	Bovinos	---	Ward *et al.,* 1997.
Lisozima	Primates	---	Messier y Stewart, 1997.
Toxinas			
Conotoxina	Gástropodo Conos	Ina	Duda y Palumbi, 1999.
Fosfolipasa A2	Serpientes crotalinas	Miyata y Yasunaga	Nakashima *et al.,* 1995.

(Continúa)

Tabla 3. Genes involucrados en actividades celulares de varias clases que se encuentran bajo selección positiva según diferentes métodos moleculares *(continúa)*

Gen y actividad celular	Organismo	Método	Referencia
Transporte de electrones y síntesis de ATP			
Subunidad F0 de la ATP sintetasa	*E. coli*	Nei-Gojobori	Endo *et al.,* 1996.
Isoforma COX7A	Primates	---	Schmidt et al., 1999.
COX4	Primates	---	Wu et al., 1997.
Citoquinas			
SF de granulocito-macrófago	Roedores	---	Shields et al., 1996.
Interleucina-3	Primates	---	Shields et al., 1996.
Interleucina-4	Primates	---	Shields et al., 1996.
Misceláneos			
CDC6	*Saccharomyces cerevisiae*	Nei-Gojobori	Endo *et al.,* 1996.
Hormona de crecimiento	Vertebrados	Nei-Gojobori	Wallis, 1996.
Cadena β de la hemoglobina	Peces antárticos	ML	Bargelloni *et al.,* 1998.
Gen Jingwei	*Drosophila*	---	Long y Langley, 1993.
Péptido C3 de la prostateína	Rata	Nei-Gojobori	Endo *et al.,* 1996.

*ML: Método de Máxima Verosimilitud

La selección a nivel genómico

La gran cantidad de secuencias almacenadas en las bases de datos (GenBank, EMBL, entre algunas) representan tanto un reto como una oportunidad para la genética de poblaciones y la genómica en cuanto al desciframiento de la historia evolutiva que contienen, así como para entender los mecanismos evolutivos que han moldeado la diversidad molecular observada. De la misma forma

han llevado al desarrollo de nuevos métodos de análisis que puedan manejar esta gran cantidad de datos y que nos permitan llevar a cabo interpretaciones efectivas acerca de los procesos evolutivos. Esta tarea es especialmente difícil si tomamos en consideración que muchos procesos evolutivos nos pueden llevar a resultados similares.

El advenimiento de la era de la genómica ha generado la disponibilidad de una gran cantidad de genomas completos, lo que se ha convertido en una herramienta poderosa para el estudio de la actuación de las distintas fuerzas evolutivas a gran escala. Esto hace que podamos hacer comparaciones de alta resolución y poder estadístico gracias al número de genes involucrados en el estudio.

Se han descrito hasta la fecha dos trabajos importantes que parecen describir los papeles que tiene la selección a nivel genómico, curiosamente ambos implican el estudio de dos dominios: el de las bacterias y el eucarionte, y de esta misma forma los resultados parecen ser antagónicos.

El primer estudio corresponde a la comparación entre las tasas de sustitución de los denominados genes esenciales y no esenciales de tres especies bacterianas: *Escherichia coli*, *Helicobacter pylori* y *Neisseria meningitidis* (tabla 4; Jordan *et al.*, 2002). En este trabajo los autores asignan la categoría de genes esenciales apoyados en los datos experimentales (sobre todo de mutantes) que existen de *E. coli,* extrapolándolos a los genes homólogos en las otras especies bacterianas. Las tasas de sustitución fueron obtenidas mediante el método de Pamilo-Bianchi-Li (Li, 1993; Pamilo y Bianchi, 1993). Mostramos a continuación la tabla 4, que es una modificación de la descrita en dicho trabajo, donde podemos ver la comparación entre las distintas tasas de sustitución sinónima (*Ks*) y no-sinónima (*Ka*) y el cociente de relación entre ambas *Ka/Ks* (equivalente a *dN/dS*).

Como podemos observar, parece existir una diferencia significativa entre los genes esenciales y no esenciales dentro del genoma de *E. coli*. De igual manera observamos que este patrón se repite para *H. pylori* y *N. meningitidis*. Además, al parecer las tasas de sustitución son mayores en éstos dos últimos patógenos en comparación con las observadas para *E. coli*, lo que parece indicar que estos organismos evolucionan más rápidamente (tabla 4). Uno de los resultados más importantes de este estudio es el hecho de que los autores no detectan selección positiva en todo el análisis, aunque no descartan la posibilidad de que en algunas regiones específicas de algunos genes exista este tipo de selección. Concluyen que, en el caso particular de las bacterias, la selección negativa o purificadora es casi la regla en la evolución a nivel

Tabla 4. Comparación de las distintas tasas de sustitución de nucleótidos y el cociente (*Ka/Ks*) entre genes esenciales y no esenciales de tres especies bacterianas

	Ks (±se)	Ka (±se)	Ka/Ks (±se)
Escherichia coli			
Genes esenciales	0.02699 ±0.0021	0.00111 ±0.0010	0.0450 ± 0.007
No esenciales	0.051 ±0.0010	0.00360 ±0.0002	0.0840 ±0.002
Helicobacter pylori			
Genes esenciales	0.11133 ±0.0041	0.01289 ±0.0011	0.1132 ± 0.009
No esenciales	0.03524 ±0.0031	0.02164 ±0.0014	0.1614 ±0.008
Neisseria meningitidis			
Genes esenciales	0.06537 ±0.0069	0.00476 ±0.0007	0.0732 ± 0.010
No esenciales	0.09156 ±0.0059	0.00960 ±0.0009	0.1765 ± 0.019

proteínas. Es decir, que la tasa evolutiva está determinada por la proporción de sitios que poseen altos coeficientes selectivos contra las mutaciones deletéreas (Jordan *et al*., 2002). En otras palabras, estos resultados sugieren que la selección purificadora es la principal fuerza moldeando la historia bacteriana a nivel genómico.

El segundo trabajo analiza el papel que juega la adaptación a nivel genómico dentro del género *Drosophila* (Smith y Eyre-Walker, 2002). Se basan en la determinación de la proporción de aminoácidos que se sustituyen debido a selección positiva durante la divergencia de las especies *D. simulans* y *D. yakuba* utilizando todos los genes homólogos existentes para ambas especies. Los valores de las sustituciones sinónimas y no-sinónimas son calculados a partir del modelo de máxima verosimilitud desarrollado por Yang y colaboradores (Goldman y Yang, 1994; Nielsen y Yang, 1998; Yang y Nielsen, 2000; Yang y Bielawski, 2000), que ya han sido descritos ampliamente. Se correlacionan estos parámetros con el polimorfismo de tipo sinónimo o no sinónimo a lo largo de la secuencia de los distintos genes. Las conclusiones de este trabajo muestran que aproximadamente un 24% de las sustituciones de aminoácidos entre ambas especies de mosca de la fruta son producto de la selección positiva. Finalmente, determinan que existen aproximadamente 270,000 sustituciones de aminoácidos seleccionadas positivamente durante la divergencia de las dos especies. Esto implica (tomando el tiempo de divergencia entre ambas que es de 6 millones de años) que estas dos especies han sufrido en promedio una sustitución del tipo adaptativo cada 45 años ó cada 450 generaciones (si contamos 10 generaciones por año), lo que indicaría que para el caso de

este grupo de eucariontes la selección natural de tipo darwiniano (positiva) posee el papel principal dentro de la evolución molecular, a diferencia de su contraparte procarionte.

Programas computacionales para determinación de selección positiva

Anteriormente se presentó una revisión de los métodos y de los resultados directos del estudio de la participación de la selección a nivel molecular, pero en la práctica ¿cómo determinamos el cociente *dN/dS* y cómo aplicamos las distintas pruebas estadísticas para probar que existe selección positiva en una muestra de secuencias de nucleótidos o amino ácidos? y ¿cómo realizamos reconstrucciones filogenéticas?

Existen varios paquetes de programas computacionales cuyo objetivo principal es el análisis y la determinación de distintos parámetros evolutivos, y que incluyen también la mayoría de las pruebas estadísticas necesarias en cualquier estudio de genética. A continuación describo cinco de los principales paquetes utilizados en la mayoría de los trabajos que descritos en el presente capítulo y en la literatura científica.

PAML (Phylogenetic Analysis by Maximum Likelihood)

Su autor es Ziheng Yang y se encuentra accesible de forma gratuita por la red en la siguiente dirección : abacus.gene.ucl.ac.uk/software/paml.html. La versión actual es la 3.13 (agosto de 2002). Puede ser instalado en computadoras con sistema operativo Windows 95/98/2000/NT , PowerMac y Unix. La mayoría de los trabajos de estudio de selección positiva usan este paquete, que incluye los siguientes programas y tipos de análisis:

- Baseml. Análisis bajo el modelo de máxima verosimilitud, cálculo de topologías de árboles filogenéticos, cálculo de largo de las ramas y parámetros de sustitución bajo varios modelos evolutivos (Jukes y Cantor, Kimura-2, etc.), análisis de reloj molecular, reconstrucción de secuencias ancestrales, etc.
- Basemlg. Lo mismo que el anterior sólo que incluye el modelo de distribución del parámetro gamma.
- Codonml. Análisis de máxima verosimilitud de secuencias codificantes usando modelos de sustitución de codones (Goldman y Yang, 1994), uso de

codones, cálculo de tasas de sustitución sinónimas y no-sinónimas, prueba estadística de selección positiva basado en cociente *dN/dS*, reconstrucción de secuencias de codones ancestrales.
- aaml. Análisis de máxima verosimilitud de aminoácidos bajo varios modelos (Poisson, Dayhoff, etc), cálculo de reloj molecular, parámetro gamma, y reconstrucción de secuencias de aminoácidos ancestrales.
- Pamp. Análisis basados en parsimonia.
- Mcmctree. Estimación de filogenias de secuencias de ADN por método Bayesiano, cálculo de probabilidades posteriores por método de Monte Carlo.
- Yn00. Método descrito por Yang y Nielsen (2000), incluye Nei-Gojobori (1986).

HYPHY (Hypothesis testing using phylogenies)

Fue desarrollado por Sergei Kosakovski y Spencer Muse, sigue siendo la versión piloto, los autores no han liberado una versión formal final, pero se puede obtener de forma gratuita desde la siguiente dirección: 222.hyphy.org. La versión es la .95 beta. Se puede instalar en computadoras con sistema operativo Windows 95/98/2000/NT, MacOX y Unix. Algunos de los análisis que incluye son: análisis de uso de codones, frecuencia de nucleótidos y aminoácidos, análisis de agrupamiento, comparación entre distintos modelos con parámetros W diferentes basados en los trabajos de Yang *et al.* (Goldman y Yang, 1994; Nielsen y Yang, 1998; Yang y Nielsen, 2000; Yang y Bielawski, 2000), análisis de reloj molecular, descomposiciones de estrella, análisis de topología.

Este paquete de programas es muy accesible a los usuarios, con ventanas muy claras y un tutorial muy completo. Por otra parte es una forma más sencilla de aproximarse a los mismos modelos de Yang et al. (Goldman y Yang, 1994; Nielsen y Yang, 1998; Yang y Nielsen, 2000; Yang y Bielawski, 2000) que en el programa de PAML parecen un poco más complicados y tienen menos instrucciones sobre su uso.

MEGA (Molecular Evolutionary Genetics analysis)

Desarrollado por Sudhir Kumar (Kumar *et al.*, 2000). Se puede bajar de forma gratuita a través de la siguiente dirección: www.megasoftware.org. La versión actual es la 2.0. Se puede instalar en computadoras con sistema operativo Windows 95/98/2000/NT, para Macintosh sólo si tiene un emulador de PC y

para SUN Workstation con Softwindows 95. No posee un límite de tamaño y número de secuencias, éste estará dado por la capacidad de la computadora, lo que lo hace muy adecuado incluso para el análisis de genomas completos.

Una gran innovación de este programa es que posee un editor de textos que es muy fácil de usar, sin límite al número y tamaño de secuencias y que nos permite agregar la extensión necesaria para el uso de otros formatos para otros programas, cualquier formato es muy fácil de convertir en MEGA pues tiene un transformador integrado. Es bastante completo en cuanto al tipo de análisis que desarrolla, incluye: cálculo de distancias genéticas, construcción de genealogías bajo todos los modelos evolutivos existentes, incluye dos pruebas de selección, una de neutralidad (Tajima D), determinación de contenidos de GC, frecuencias de aminoácidos y uso de codones, determinación de tasas de sustitución por método de Nei-Gojobori y método modificado de Kumar.

DNASP

Desarrollado por Rozas y Rozas (1999), este es uno de los mejores paquetes; su única limitante es el número de nucleótidos que puede analizar, aproximadamente 16,000 pares de bases, que para fines prácticos es bastante amplio pero no muy adecuado para estudios a nivel genómico. La versión actual es la 3.98.6. Incluye los siguientes análisis: varias pruebas de neutralidad como Hudson-Kreitman-Aguadè (HKA), Tajima, McDonald y Kreitman; análisis de coalescencia, polimorfismo de ADN, varias pruebas de uso de codones, determinación de tasas de sustitución por el método de Nei-Gojobori, flujo génico, desequilibrio de ligamiento y recombinación.

Se puede acceder de manera gratuita en la red a través de la siguiente dirección: www.ub.es/DNAsp/ y se puede instalar para sistema operativo Windows 95/98/2000/NT y en Macintosh si tiene un simulador de PC como Softwindows.

PHYLIP (Phylogeny inference package)

Creado por Joe Felsenstein (Felsenstein, 1993), se puede bajar en el siguiente sitio: //evolution.genetics.washington.edu/phylip.html/. Este paquete es uno de los pioneros en el análisis de evolución molecular y construcción filogenética, que además de incluir métodos de distancia como el vecino más cercano (NJ) y Máxima Verosimilitud, incluye el método de parsimonia.

El autor fue el primero en aplicar máxima verosimilitud al análisis genético y casi todos los trabajos que lo utilizan derivan de su método. Es un buen sitio para consultar qué otros programas de análisis evolutivo existen dependiendo de qué modelos se quieren utilizar. El autor lo mantiene muy actualizado, es bastante descriptivo y uno de los más utilizados en cualquier estudio filogenético. Se puede instalar en casi cualquier plataforma Windows en todas sus versiones, Macintosh y UNIX.

Bibliografía

Baric, R.S. *et al.* 1997. Episodic evolution mediates interspecific transfer of a murine coronavirus. *Journal of Virology* 71:1946-1955.

Bargelloni, L. *et al.* 1998. Antarctic fish hemoglobins: evidence for adaptive evolution at subzero temperatures. *PNAS* 95:8670-8675.

Bishop, J.A. y L.M. Cook. 1975. Moth, melanism and clean air. *Scientifi American* 232(1):90-99.

Bishop, J.A. *et al.* 2000. Rapid evolution in plant chitinases: molecular targets of selection in plant-pathogen coevolution. *PNAS* 97:5322-5327.

Cameron, J.M. 1995. A method for estimating the numbers of synonymous and non synonymous substitutions per site. *Journal of Molecular Evolution* 41:1152-1159.

Duda, T.F. y S.R. Palumbi. 1999. Molecular genetics of ecological diversification: duplication and rapid evolution of toxin genes of the venomous gastropod *Conus*. *PNAS* 96:6820-6823.

Efron, B. y R.J. Tibshirani. 1993. *An introduction to the bootstrap*. Chapman y Hall, New York.

Endo, T. *et al.* 1996. Large-scale search for genes on which positive selection may operate. *Molecular Biology Evolution* 13:685-690.

Felsenstein, J. 1981. Evolutionary trees from ADN sequences: a maximum likelihood approach. *Journal of Molecular Evolution* 17:368-376.

———. 1993. Phylogeny inference package (PHYLIP).Version 3.5. University of Washington, Seattle.

Fitch, W.M. *et al.* 1997. Long term trends in the evolution of H(3) HA1 human influenza type A. *PNAS* 94:7712-7718.

Ford, M.J. *et al.* 1999. Natural selection promotes divergence of transferrin among salmonid species. *Molecular Ecology* 8:1055-1061.

Fu, Y.X. y W.H. Li. 1993. Statistical tests of neutrality mutations. *Genetics* 133:693-709.

Futuyma, D. 1986. *Evolutionary biology*. 2a. ed. Sinahuer Associates Inc. Sunderland. Massachusetts.

Goodwin, R.L. *et al.* 1996. Patterns of divergence during evolution of a_1- proteinase inhibitors in mammals. *Molecular Biology and Evolution* 13:346-358.

Goldman, N. y Z. Yang. 1994. A codon-based model of nucleotide substitution for protein coding ADN sequences. *Molecular Biology and Evolution* 11(5):725-736.

Grant, P.R. 1986. *Ecology and evolution of Darwin's finches*. Princeton University Press, Princeton, New Jersey.

Haydon, D.T., A.D. Bastos, N.J. Knowles y A.R. Samuel. 2001. Evidence for positive selection in Foot-and-Mouth disease virus capsid genes from field isolates. *Genetics* 157:7-15.

Hudson, R.R., M. Kreitman y M. Aguadè. 1987. A test of neutral molecular evolution based on nucleotide data. *Genetics* 116:153-159.

Hughes, A.L. 1992. Positive selection and interallelic recombination at the merozoite surface antigen-1 (MSA-1) locus of *Plasmodium falciparum*. *Molecular Biology Evolution* 9:381-393.

———. 1995. The evolution of the type I interferon family in mammals. *J Mol Evol* 41:539-548.

Hughes, A.L. y M. Nei. 1988. Pattern of nucleotide subtitution at major histocompatibility complex class I loci reveals overdominant selection. *Nature* 335:167-170.

Hughes, A.L. y M. Yeager. 1997. Coordinated amino acid changes in the evolution of mammalian defensines. *Journal of Molecular Evolution* 44:675-682.

Hughes, M.K. y A.L. Hughes. 1995. Natural selection on *Plasmodium* surface proteins. *Molecular and Biochemical Parasitology* 71:99- 113.

Ina, Y. 1995. New methods for estimating the numbers of synonymous and non synonymous substitutions. *Journal of Molecular Evolution* 40:190-226.

Ishimizu, T. *et al.* 1998. Identification of regions in which positive selection may operate in S-Rnase of Rosaceae: implications for S-allele-specific recognition sites in S-Rnase. *FEBS Lett* 440:337-342.

Jope, E.M. 1976. The evolution of plants and animals under domestication: the contribution of studies at the molecular level. *Philosophical Transactions Royal Soc iety of London Series B, Biological Sciences* 275 (936):99-116.

Jordan, I.K., I.B. Rogozin, Y.I. Wolf y E.V. Koonin. 2002. Essential genes are more evolutionarily conserved than are nonessential genes in bacteria. *Gen Res* 12:962-68.

Jukes, T.H. y C.R. Cantor. 1969. *Evolution of protein molecules*. En: H.N. Munro (ed.), *Mammalian protein metabolism.* Pp 21-132. Academic Press, New York.

Karn, R.C. y M.W. Nachman. 1999. Reduced nucleotide variability at an androgen-binding protein locus (Abpa) in house mice: evidence for positive natural selection. *Molecular Biology and Evolution* 16:1192-1197.

Kimura, M. 1968. Evolutionary rate at the molecular level. *Nature* 217(129):624-626.

———. 1980. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *Journal of Molecular Evolution* 16:111-120.

———. . 1983. *The neutral theory of molecular evolution*. Cambridge University Press. Cambridge.

Kitano, T. *et al.* 1998. Conserved evolution of the Rh50 gene compared to its homologous Rh blood group gene. *Biochemical and Biophysical Research Communications* 249:78-85.

Kumar, S., K. Tamura, I. Jakobsen y M. Nei. 2000. *MEGA*: Molecular evolutionary genetics analysis, ver 2. Pennsylvania State University, University Park, and Arizona State University, Tempe.

Lack, D. 1947. *Darwin's finches.* Cambridge, University Press, Massachusetts.

Li, W.H. 1997. *Molecular Evolution*. Sinauer Publishers, Massachusetts.

———. 1993. Unbiased estimation of the rates of synonymous and nonsynonymous substitution. *J Mol Evol* 36:96-99.

Li, W.H., C.I. Wu y C.C. Luo. 1985. A new method for estimating synonymous and non synonymous rates of nucleotide substitution considering the relative likelihood of nucleotide and codon changes. *Molecular Biology and Evolution* 2:150-174.

Long, M. y C.H. Langley. 1993. Natural selection and the origin of *jigwei*, a chimeric processed functional gene in *Drosophila*. *Science* 260:91-95.

McDonald, J. y M. Kreitman. 1991. Adaptive protein evolution at *adh* locus in *Drosophila*. *Nature* 351:652-654.

Messier, W. y C.B. Stewart 1997. Episodic adaptive evolution of the primate lysozymes. *Nature* 385:151-154.

Miyata, T. y T. Yasunaga. 1980. Molecular evolution of mRNA: a method for estimating evolutionary rates of synonymous and amino acid substitutions from homologous nucleotide sequences and its application. *Journal of Molecular Evolution* 16:23-36.

Muse, S.V. y B.S. Gaut. 1994. A likelihood approach for comparing synonymous and nonsynonymous nucleotide substitution rates, with application to the chloroplast genome. *Molecular Biology and Evolution* 11(5):715-724.

Nakashima, K. *et al.* 1995. Accelerated evolution in the protein-coding regions is universal in crotaline snake venom gland phospholipase A_2 isozyme genes. *PNAS* 92:5605-5609.

Nei, M. y T. Gojobori. 1986. Simple methods for estimating the numbers of synonymous and non synonymous nucleotide substitutions. *Molecular Biology and Evolution* 3:418-426.

Nei, M. y S. Kumar. 2000. *Molecular evolution and phylogenetics*. Oxford University Press, New York.

Nei, M. y WH. Li. Mathematical model for studying genetic variation in terms of restriction endonucleases. *PNAS* 76:5269-5273.

Nielsen, R. y Z. Yang. 1998. Likelihood models for detecting positively selected amino acid sites and applications to the HIV-1 envelope gene. *Genetics* 148(3):929-936.

Ohta, T. 1992. The nearly neutral theory of molecular evolution. *Ann Rev Ecol Syst* 23:263-286.

Ota, T. y M. Nei. 1994. Variances and covariances of the number of synonymous and non-synonymous substitutions per site. *Molecular Biology and Evolution* 11:613-619.

Otto, S.P. 2000. Detecting the form of selection from ADN sequence data. *TIG* 16(12):526-529.

Pamilo, P. y N.O. Bianchi. 1993. Evolution of the *Zfx* and *Zfy* genes:rates and interdependence between the genes. *Molecular Biology and Evolution* 10:271-281.

Pamilo, P. y R.W. O'Neill. 1997. Evolution of *Sry* genes. *Molecular Biology and Evolution* 14:49-50.

Riley, M.A. 1993. Positive selection for colicin diversity in bacteria. *Molecular and Biology Evolution* 10:1048-1059.

Rooney, A.P. y J. Zhang. 1999. Rapid evolution of a primate sperm protein: relaxation of functional constraint or positive darwinian selection? *Molecular and Biology Evolution* 16:706-710.

Sato, A., C. O'Huigin, F. Figueroa, P.R. Grant, B.R. Grant, H. Tichy y J. Klein. 1999. Phylogeny of Darwin's finches as revealed by mtADN sequences. *PNAS* 96(9):5101-5106.

Schmidt, T.R. *et al.* 1999. Molecular evolution of the COX7A gene family in primates. *Mol Biol Evol* 16:619-626.

Shields, D.C. *et al.* 1996. Evolution of hemopoietic ligands and their receptors: influence of positive selection on correlated replacements throughout ligand and receptor proteins. *J Immunol* 156:1062-1070.

Simmonsen, K.L., G.A. Churchill y C.F. Aquadro. 1995. Properties of statistical tests of neutrality for ADN polymorphism data. *Genetics* 141:413-429.

Smith, N.H. *et al.* 1995. Sequence evolution of the porB gene of *Neisseria gonorrhoeae* y *Neisseria meningitidis*: evidence for positive darwinian selection. *Molecular and Biology Evolution* 12:363-370.

Smith, N.G.C. y A. Eyre-Walker. 2002. Adaptive protein evolution in *Drosophila*. *Nature* 415:1022-24.

Stotz, H.U. *et al.* 2000. Identification of target amino acids that affect interactions of fungal polygalacturonases and their plant inhibitors. *Physiological and Molecular Plant Pathology* 56:117-130.

Sutton, K.A. y M.F. Wilkinson. 1997. Rapid evolution of a homeodomain: evidence for positive selection. *Journal of Molecular Evolution* 45:579-588.

Suzuki, Y. y T. Gojobori. 1999. A method for detecting positive selection at single amino acid sites. *Molecular and Biological Evolution* 16:1315-1328.

Tajima, F. 1983. Evolutionary relationship of ADN sequences in finite populations. *Genetics* 105:437-460.

———. 1989. Statistical method for testing the neutral mutation hypothesis by ADN polymorphism. *Genetics* 123:585-595.

Tanaka, T. y M. Nei. 1989. Positive darwinian selection observed at the variable-region genes of immunoglobulins. *Molecular and Biological Evolution* 6:447-459.

Ting, C.T. *et al.* 1998. A rapidly evolving homeobox at the site of a hybrid sterility gene. *Science* 282:1501-1504.

Tsaur, S.C. y Wu, C.I. 1997. Positive selection and the molecular evolution of a gene of male reproduction, Acp26Aa of *Drosophila*. *Molecular and Biological Evolution* 14:544-549.

Vacquier, V.D. *et al.* 1997. Positive darwinian selection on two homologous fertilization proteins: what is the selective pressure driving their divergence? *J Mol Evol* 44:15-22.

Wallis, M. 1996. The molecular evolution of vertebrate growth hormones: a pattern of near-stasis interrupted by sustained burst of rapid change. *Journal of Molecular Evolution* 43:93-100.

Ward, T.J. *et al.* 1997. Nucleotide sequence evolution at the k.casein locus: evidence for positive selection within the family Bovidae. *Genetics* 147:1863-1872.

Watterson, G.A. 1975. On the number of segregating sites in genetical models without recombination. *Theoretical Population Biology* 7:256-276.

Wiesenfeld, S.L. 1967. Sickle-cell trait in human biological and cultural evolution. Development of agriculture causing increased malaria is bound to gene-pool changes causing malaria evolution. *Science* 157 (793):1134-1140.

Wu, J.C. *et al.* 1999. Recombination of hepatitis D virus RNA sequences and its implications. *Molecular Biology and Evolution* 16:1622-1632.

Wu, W. *et al.* 1997. Molecular evolution of cytochrome c oxidase subunit IV: evidence for positive selection in simian primates. *Journal of Molecular Evolution* 44:477-491.

Yamaguchi-Kabata, Y. y T. Gojobori. 2000. Reevaluation of amino acid variability of the human immunodeficiency virus type 1 gp120 envelope glycoprotein

and prediction of new discontinuous epitopes. *Journal of Virology* 74:4335-4350.

Yang, Z. y J. Bielawski. 2000. Statistical methods for detecting molecular adaptation. *TREE* 15(12):496-503.

Yang, Z., S. Kumar y M. Nei. 1995. A new method of inference of ancestral nucleotide and amino acid sequences. *Genetics* 141(4):1641-50.

Yang, Z. y R. Nielsen. 2000. Estimating synonymous and non synonymous substitution rates under realistic evolutionary models. *Molecular Biology and Evolution* 17:32-43.

Zhang, J. y M. Nei. 1997. Accuracies of ancestral amino acid sequences inferred by the parsimony, likelihood and distance methods. *Journal of Molecular Evolution* 44 (Suppl 1):S139-46.

Zhang, J., S. Kumar y M. Nei. 1997. Small-sample test of episodic adaptive evolution: a case study of primate lysozymes. *Molecular Biology and Evolution* 14:1335-38.

Zhang, J. *et al.* 1998. Positive Darwinian selection after gene duplication in primate ribonuclease genes. *PNAS* 95:3708-3713.

Zhang, Y.X. *et al.* 2002. Genome shuffling leads to rapid phenotypic improvement in bacteria. *Nature* 415 (6872):644-646.

Zuckerkandl, E. y L. Pauling. 1965. Molecules as documents of evolutionary biology. *Journal of Theoretical Biology* 8(2):357-366.

Capítulo 2

Flujo génico: métodos para estimarlo y marcadores moleculares

Erika Aguirre Planter

El flujo génico se refiere a todos los mecanismos que generan movimiento de genes de una población a otra. Las poblaciones de una especie pueden intercambiar genes en mayor o menor grado, ya sea genes nucleares o genomas uniparentales como la mitocondria o el cloroplasto, debido al movimiento de gametos, semillas, individuos juveniles o adultos (en animales y en el caso de algunas plantas), así como por eventos de extinción y recolonización de poblaciones enteras (Slatkin, 1985a), aunque cabe mencionar que los migrantes que no se reproducen en la población a la que migraron no contribuyen al flujo génico.

El estudio del flujo génico ha sido un tema vital en biología evolutiva, ya que es un componente importante en la estructura de las poblaciones. Anteriormente se pensaba que el flujo génico en general era muy restringido y de poca importancia evolutiva (Levin, 1981). Sin embargo, se ha observado en distintas especies que los niveles estimados de flujo génico generalmente son altos (Riesberg y Burke, 2001) y que pueden actuar como una fuerza que mantiene integrada a la especie, además de influir en procesos ecológicos. Por ejemplo pueden determinar la persistencia y adaptación de poblaciones locales, las tasas de extinción de las especies, la evolución de los rangos de distribución de las especies y otras propiedades ecológicas (Whitlock y McCauley, 1999). Asimismo, si el flujo génico entre poblaciones de una especie es alto, todas ellas evolucionan de manera conjunta (a menos que sea contrarrestado por

selección o deriva génica), pero si es muy bajo, empiezan a divergir, lo que puede contribuir al aislamiento reproductivo y al establecimiento de linajes evolutivamente independientes (Slatkin, 1994).

En biología de la conservación se ha tratado de estimar el flujo génico actual tratando de entender el movimiento de genes a una escala regional o de paisaje. Por ejemplo, existen estudios que consideran que las poblaciones fragmentadas pueden tener una dinámica de metapoblaciones a través de la extinción y recolonización de distintos fragmentos, aunque actualmente no se tiene una metodología clara en la que las estimaciones de flujo génico permitan relacionar los modelos de metapoblaciones con la ecología del paisaje.

Para tratar de medir estas tasas de flujo génico se han utilizado métodos directos e indirectos. En este capítulo abordaremos con especial énfasis los indirectos, ya que son los que se basan en datos moleculares.

Métodos directos

Se basan en observaciones o experimentos que miden el grado de dispersión de gametos o individuos, por ejemplo con la captura y recaptura de individuos marcados. Puede también hacerse un análisis de parentesco para identificar padres y después cuantificar el patrón del movimiento de genes. Los análisis de parentesco nos permiten estimar la distribución de las distancias de dispersión y examinar el movimiento de genes por polen y semilla. Existen distintos modelos para el análisis de parentesco, pero el modelo general utiliza la progenie de madres conocidas para asignar la paternidad a un grupo de donadores o padres potenciales y así cuantificar el movimiento de genes (véanse los capítulos 5 y 6 de este libro).

En este sentido, Meagher (1986) presentó un análisis de paternidad en una población de plantas en el que cuantificó la varianza en el éxito reproductivo como una función de la distancia. Modificaciones subsecuentes a este método han sido utilizadas para el estudio del movimiento de genes en poblaciones (Devlin y Ellstrand, 1990; Roeder *et al.*, 1989; Smouse y Meagher, 1994), aunque la estimación directa de flujo génico que dan se basa generalmente en uno o dos episodios reproductivos.

A raíz del uso de los métodos directos, se concluye que en la mayoría de las especies la dispersión es muy localizada, aunque esto no quiere decir necesariamente que el flujo génico en poblaciones sea muy restringido. Esto se debe a que se subestima la frecuencia de la dispersión a larga distancia, que puede ser difícil de detectar, no se toman las extinciones y recolonizaciones

como una fuente de flujo génico y no se detectan eventos raros que pueden ser importantes (Slatkin, 1985a). Asimismo, las estimaciones de flujo génico obtenidas con este tipo de datos no pueden ser interpretadas como flujo génico entre poblaciones, ya que miden la migración en un área determinada que no es necesariamente una población, y en un periodo dado, por lo que no puede ser considerado flujo génico en una escala evolutiva. De la misma manera, las medidas directas de dispersión no necesariamente reflejan el movimiento de genes, porque no se sabe si el migrante se reproduce exitosamente (Whitlock y McCauley, 1999).

Sin embargo, al reflejar el flujo génico instantáneo, los métodos directos aportan detalles, como la etapa de la historia de vida en la que la dispersión ocurre, que nos dan una idea de las condiciones ecológicas favorables para la dispersión. Por ello, estos métodos son apropiados para el estudio del movimiento de genes en una escala ecológica o fina.

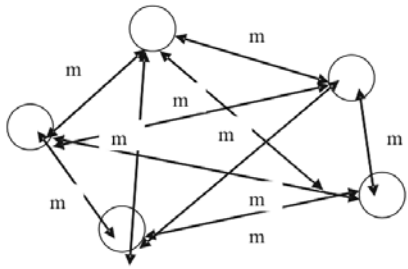
Métodos indirectos

Gracias al desarrollo de técnicas y análisis moleculares, es posible estimar el flujo génico de una manera detallada y con mayor resolución. Se basan principalmente en observar la distribución espacial de alelos en las poblaciones para hacer inferencias de los niveles o patrones de flujo génico en las poblaciones (Slatkin, 1985a).

La mayoría de los modelos teóricos de flujo génico surgen de los conceptos desarrollados por Sewall Wright y suponen que los organismos están formando poblaciones discretas -modelo de islas- que se diferencian por mutación y deriva génica (Wright, 1943) o bien que forman poblaciones con una distribución continua -modelos de aislamiento por distancia- en donde la probabilidad de flujo génico disminuye al incrementarse la distancia espacial.

El modelo usado comúnmente para estimar flujo génico es el modelo de islas infinitas (*infinite islands model*) de Wright (1951; figura 1). Este modelo considera condiciones en equilibrio entre un número infinito de islas o subpoblaciones de igual tamaño, que intercambian migrantes entre cualquiera de las islas con igual probabilidad, a una tasa constante. Las poblaciones pueden ser tratadas como réplicas, y el modelo se puede caracterizar con sólo dos parámetros: tamaño poblacional (N) y tasa de migración (m). La importancia de la deriva génica es proporcional a $1/N$, mientras que la importancia del flujo génico es proporcional a m (Slatkin, 1985a).

Figura 1. Modelo de islas infinitas, donde *m* es la tasa de migración



Wright (1951) introdujo un método para dividir el coeficiente de endogamia en una población subdividida en un componente debido a apareamientos no aleatorios dentro de poblaciones (F_{is}) y otro componente debido a la subdivisión entre poblaciones (F_{st}). F_{st} es definido como:

$$F_{st} = (H_T - H_S)/H_T$$

donde H_T es el promedio de la heterocigosis esperada en la población total, para todos los loci, y H_S es el promedio de la heterocigosis esperada dentro de subpoblaciones para todos los loci.

F_{st} mide la variación de las frecuencias alélicas entre poblaciones, y por tanto la diferenciación genética entre ellas. Se relaciona con la migración de la siguiente manera: la tasa a la que un alelo se fija en una población es inversamente proporcional al tamaño efectivo de la población, N_e (N por simplicidad). Esta tasa de fijación puede ser contrarrestada por el flujo génico que llega de otras poblaciones a una tasa m. Estos factores llegan a un equilibrio en donde F_{st} es aproximadamente:

$F_{st} \approx 1/ (4\,Nm + 1)$ para genomas diploides
$F_{st} \approx 1/ (2\,Nm + 1)$ para genomas haploides

F_{st} puede estimarse mediante aloenzimas o datos moleculares de ADN y el flujo génico efectivo puede posteriormente calcularse en el equilibrio de manera teórica como:

$Nm \approx ¼ (1/F_{st} - 1)$ para organismos diploides
$Nm \approx ½ (1/F_{st} - 1)$ para organismos haploides

Bajo este modelo, de acuerdo con Wright (1969), una tasa de migración >1 en cada generación es suficiente para contrarrestar la diferenciación genética debida a deriva génica.

Una medida similar es la propuesta por Crow y Aoki (1984), en la que existe un número finito de islas (modelo de *n*-islas), en el que el equivalente para alelos múltiples de F_{st} es el parámetro G_{st} (Nei, 1973) definido como:

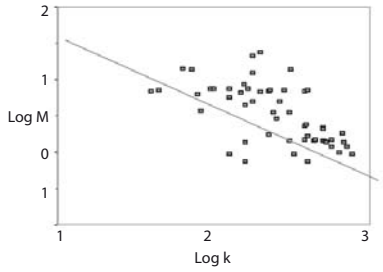
$$G_{st} \approx 1/ (4Nm\alpha + 1)$$

Donde $\alpha = [n/ (n\text{-}1)]^2$ y n es el número de subpoblaciones

Este modelo asume que las poblaciones están en equilibrio y que las islas intercambian migrantes entre sí con igual probabilidad a una tasa constante.

Otro modelo relacionado con el flujo génico, como se mencionó anteriormente, es el de aislamiento por distancia, que examina la relación del flujo génico entre pares de poblaciones (*M*, equivalente a *Nm* entre pares de poblaciones) con la distancia geográfica (Slatkin, 1993). Bajo este modelo, los valores de *M* se estiman a partir de valores de θ entre pares de poblaciones (que es una medida de diferenciación entre poblaciones, Weir y Cockerham 1984) y se aplica una regresión de log10*M* en log10*k* (*k*= distancia geográfica). Para determinar si existe una relación lineal significativa entre el flujo génico y la distancia geográfica entre poblaciones se utiliza comúnmente una prueba de Mantel (1967), ya que los valores de *M* para los diferentes pares de poblaciones no son independientes (figura 2).

Figura 2. Gráfica entre *k* = distancia geográfica entre pares de poblaciones y *M* = flujo génico entre pares de poblaciones



Otro método para estimar *Nm* es el de alelos privados (Slatkin 1981, 1985b), en el que con base en muestras de distintas localidades se obtiene $p(1)$, que es la frecuencia promedio de alelos que sólo están en una muestra. Slatkin (1985b) encontró que para el modelo de islas y el de *stepping stone* $\log_{10}[p(1)]$ está linealmente relacionado con $\log_{10}(Nm)$, por lo que:

$$\log10[p(1)] = a\log10(Nm) + b$$

donde $p(1)$ es la frecuencia promedio de los alelos privados y a y b son constantes que dependen del número de individuos muestreados en cada subpoblación y son determinadas por datos simulados (ver Slatkin, 1985b). Este método y la estimación de *Nm* a partir de F_{st} son parecidos, al basarse en la distribución de las frecuencias alélicas entre poblaciones. Sin embargo, simulaciones han demostrado que el método de alelos privados es más sensible a errores en la colección de datos (Slatkin, 1994).

En muchos aspectos F_{st} es un parámetro ideal que nos da una idea de la historia de las poblaciones estudiadas, dando información acerca de la importancia evolutiva del flujo génico y la deriva génica. Es relativamente fácil obtener los datos necesarios para calcular F_{st} y luego *Nm*. Sin embargo, una crítica común a los estimadores de flujo génico como F_{st} que utilizan la variación en las frecuencias alélicas, es que asumen un modelo poco real que considera un número infinito de poblaciones, donde hay ausencia de selección o mutación, que el flujo génico no es afectado por la distancia geográfica entre poblaciones y que cada población está en un equilibrio entre migración y deriva (Niegel, 2002).

Slatkin y Barton (1989) revisaron la relación entre F_{st} y flujo génico, tanto teóricamente como mediante simulaciones y encontraron evidencia para apoyar el uso del modelo de islas. Concluyeron que en una población subdividida que está en equilibrio demográfico, tanto el método de F_{st} como el de alelos privados proveen estimados razonablemente correctos de *Nm* bajo una variedad de condiciones.

Por otro lado, Whitlock y McCauley (1999) señalan que existe evidencia de que muchos de los loci a partir de los cuales se estima F_{st} están sujetos a selección, que puede tanto sobreestimar (si la selección favorece los mismos alelos en diferentes áreas) como subestimar (si la selección favorece diferentes alelos en distintas áreas) los valores de F_{st}. Por ejemplo, cuando la tasa de migración es muy baja (y en ausencia de deriva génica), la selección puede dominar el patrón de diferenciación, ya que el ligamiento a alelos sujetos a

selección local puede aumentar la F_{st} (Whitlock y McCauley, 1999). Por otro lado, Slatkin y Barton (1989) encontraron que la F_{st} en el modelo de islas no es muy sensible a la selección que favorece a uno o dos alelos, sino que sólo es importante si diferentes alelos se encuentran favorecidos en distintas localidades geográficas. El supuesto de que los alelos sean neutros o de que la selección no esté alterando el índice puede evaluarse por el grado de consistencia entre los diferentes loci a partir de los cuales se obtiene F_{st}. Si con cada loci polimórfico se obtiene aproximadamente el mismo valor de F_{st}, es muy probable que la selección no esté determinando el valor encontrado.

La tasa de mutación también puede afectar los patrones de diferenciación genética entre poblaciones. En el caso del estimador F_{st} éste no es muy sensible a las tasas de mutación, sólo si los tamaños poblacionales son pequeños (de 1000 o menores) y las tasas de mutación son mucho menores que las tasas de migración (Crow y Aoki, 1984). En marcadores como microsatélites, las tasas de mutación pueden ser muy altas, lo que puede disminuir F_{st} y sobreestimar *Nm* (Hedrick, 1999). Para resolver este problema, Slatkin (1991) definió F_{st} en términos de tiempos de coalescencia entre alelos y desarrolló el índice R_{st}, que es relativamente insensible a tasas de mutación y más apropiado para loci que mutan en forma de paso a paso (*step-wise model*). Sin embargo, R_{st} tiene una mayor varianza que otras medidas de diferenciación entre subpoblaciones y es muy sensible a violaciones del modelo de mutación paso a paso (Gaggiotti *et al.*, 1999), por lo que no es claro si en la práctica ofrece ventajas sobre F_{st}, incluso con marcadores que teóricamente se ajustan al modelo de mutación paso a paso.

Otro problema en relación a la estimación de F_{st} , es el del equilibrio. El tiempo en generaciones requerido para que F_{st} esté cerca del equilibrio es $1/[2m+1/(2N)]$, donde m es la tasa de migración y N es el tamaño poblacional (Crow y Aoki, 1984). Cuando no hay flujo génico (m=0), el tiempo promedio requerido para el equilibrio sería $2N$. Si existieron poblaciones grandes que estuvieron históricamente en contacto y que actualmente se encuentran aisladas, estarían fijadas para distintos alelos en el equilibrio, pero el tiempo requerido para aproximarse a ese equilibrio podría ser mayor a la edad de la mayoría de estas especies. Esto también podría ocurrir si los sitios muestreados han sido recientemente colonizados y por lo tanto las poblaciones todavía no estarían en equilibrio. En estos casos, los estimados darían estimaciones de flujo génico diferentes de cero aún entre especies que han estado aisladas reproductivamente durante mucho tiempo (Niegel, 2002).

El modelo de islas es una abstracción conveniente que aísla los efectos opuestos de deriva génica y flujo génico, y aunque F_{st} no siempre da estima-

ciones "exactas", es poco probable que estos estimados estén sobrestimados por varios órdenes de magnitud (Niegel, 2002). Este mismo autor propone que el número de poblaciones no tiene que ser infinito, ni necesariamente muy grande para que F_{st} dé una estimación acertada de *Nm*, y que la mutación y la selección sólo son importantes cuando las poblaciones son muy grandes. También demuestra que aunque el flujo génico esté limitado por distancia, se espera que el valor total de F_{st} sea similar al obtenido por el modelo de islas (Slatkin y Barton, 1989). El problema del uso de F_{st} surge cuando la deriva génica es un proceso débil (los tamaños efectivos de las poblaciones son grandes). F_{st} puede pensarse como un balance entre deriva génica y flujo génico y valores muy pequeños de flujo génico (<<1) pueden acercar F_{st} a 0.

En conclusión, los datos y simulaciones indican que el uso de F_{st} para estimar *Nm* es útil y sencillo. Los métodos indirectos tienen la ventaja de poder incorporar los efectos de todos los componentes históricos de la dispersión y generar un promedio de la variación en la dispersión a través del tiempo. Finalmente, los niveles de flujo génico estimados con F_{st} reflejan flujo génico histórico, no el flujo génico que está ocurriendo en el presente.

Métodos genealógicos

La distribución de la variación genética en poblaciones está fuertemente influenciada por las fuerzas microevolutivas que están actuando, como el flujo génico y la selección, así como por la historia filogenética de las poblaciones y las especies (Schaal *et al.*, 2003). Es por ello que el uso de estadísticos tradicionales para determinar la estructura de las poblaciones, como los que se mencionaron anteriormente, no toman en cuenta las relaciones evolutivas entre los alelos y no permiten distinguir entre efectos históricos y procesos recurrentes.

Muchas secuencias de ADN cuentan con una alta variación molecular neutral que puede ser utilizada para reconstruir su genealogía. En este contexto, el uso de secuencias para inferir procesos evolutivos en el pasado a nivel poblacional, ha sido enriquecido por el concepto de filogeografía propuesto por Avise (1994; véase el capítulo 14 de este libro), que examina la distribución de la genealogía de alelos o haplotipos en un contexto geográfico, utiliza la información histórica contenida en los genes y caracteriza la subdivisión de poblaciones al reconocer patrones geográficos de la estructura genealógica a lo largo del rango de distribución de una especie.

A nivel poblacional es preferible el uso de redes de haplotipos en lugar de árboles filogenéticos, ya que la inferencia cladística de relaciones filogenéticas

requiere que las variantes genéticas no formen linajes reticulados (Schaal *et al.*, 1998). Las genealogías trazan las relaciones evolutivas entre haplotipos o alelos (Schaal *et al.*, 2003) que permiten reconstruir los procesos evolutivos con los principios de la coalescencia. Los métodos que utilizan estos principios tratan de modelar el pasado utilizando un proceso estocástico, esto es, la coalescencia, basado en el concepto de que los alelos en una población pueden ser rastreados hacia atrás en el tiempo hasta el punto en el que coalescen en un alelo ancestral común (véase el capítulo 9 de este libro).

La idea principal detrás de la coalescencia es que en ausencia de selección los linajes muestreados pueden ser vistos hacia atrás en el tiempo "escogiendo" al azar a sus padres y cuando dos linajes escogen al mismo padre hay un evento de coalescencia. Eventualmente todos los linajes coalescen en un mismo y único linaje, por ejemplo, el ancestro común más reciente (Rosenberg y Nordborg, 2002). Los métodos genealógicos son utilizados para estimar los parámetros del proceso genealógico que da como resultado a cada árbol. El proceso de coalescencia es una herramienta matemática poderosa que puede ser utilizada para estimar parámetros poblacionales como las tasas de mutación o migración y otros procesos (Rosenberg y Nordborg, 2002).

Templeton *et al.* (1987, 1995) y Templeton (1998) han propuesto un análisis basado en los principios de coalescencia para conocer las relaciones entre poblaciones, que es el análisis de clados anidados (*nested clade analysis* o NCA; ver capítulo 15). Este análisis permite obtener información de procesos históricos, como estimaciones más cercanas y más realistas de flujo génico, deriva génica o expansión, además de ayudar a estimar parámetros más complejos y modelos reales de estructura poblacional en un contexto geográfico. Avise (2000) propone que los cambios en tamaños poblacionales o en niveles de flujo génico a través del tiempo dejan patrones que pueden reconocerse y que son marcas de historias poblacionales particulares. Por ejemplo, se predice que una población que se ha estado expandiendo va a presentar una filogenia en forma de estrella en (Avise, 2000). En el análisis de Templeton *et al.* (1995), se reconstruyen redes de haplotipos que se unen en clados anidados con el programa TCS (*phylogenetic network estimation using statistical parsimony*; Clement *et al.*, 2000). Este algoritmo de anidamiento agrupa haplotipos y después clados, dependiendo del número de cambios mutacionales, separándolos paso a paso hasta que todos los haplotipos/clados quedan agrupados y unidos en grupos de uno, dos o más pasos mutacionales, donde el último nivel de anidamiento comprende la red entera (Templeton *et al.*, 1987). La relación entre haplotipos/clados y la geografía se puede exami-

nar a través de un análisis estadístico llamado *de permutaciones*, utilizando por ejemplo el programa GEODIS (*A tool to perform the NCPA;* Posada *et al.*, 2000), que permite evaluar si existen asociaciones significativas entre la ubicación geográfica y las posiciones del haplotipo o del clado. En clados en los que hay una asociación significativa, se puede hacer un análisis de distancia comparando la distancia del clado D_c, que mide el rango geográfico de un clado en particular y la distancia del clado anidado D_n, que mide cómo un clado en particular se distribuye geográficamente en relación a sus clados evolutivamente más cercanos o clados hermanos (clados en la misma categoría de anidamiento). En particular, la distancia del clado D_c mide la distancia promedio que existe entre un individuo que posee un haplotipo del clado de interés y el centro geográfico de todos los individuos que poseen haplotipos del mismo clado. La distancia del clado anidado D_n mide la distancia promedio que existe entre un individuo que posee un haplotipo del clado de interés y el centro geográfico de todos los individuos que poseen haplotipos del siguiente nivel de anidamiento, que contiene al clado de interés (Templeton, 1998). Los contrastes entre estas medidas de distancia entre clados externos (clados que no son nodos interiores en la red de haplotipos) y clados internos inmediatos son importantes para discriminar las causas potenciales de la estructuración geográfica de la variación genética (Templeton *et al.*, 1995; Templeton, 1998). Para determinar si las distintas medidas de distancia son significativas se realiza un análisis de permutaciones, el cual simula la hipótesis nula de una distribución geográfica al azar para todos los clados dentro de una categoría o nivel de anidamiento (Templeton, 1998). Los resultados se interpretan utilizando la clave de inferencias de Templeton (2004), que proporciona una serie de criterios que permiten discriminar entre el papel que han jugado los eventos de flujo génico, el aislamiento por distancia, la expansión y los eventos de colonización a larga distancia en los procesos históricos que generan las relaciones observadas entre haplotipos/clados.

Aunque estos patrones descriptivos ayudan a entender cómo usar datos de genética de poblaciones para reconstruir la historia de una especie, el problema es que muchas veces el mismo patrón puede ser explicado por diferentes historias poblacionales; por ejemplo la filogenia en forma de estrella puede deberse a una población continua que se ha fragmentado. Existen muchas estructuras o historias poblacionales que son compatibles con un conjunto de datos particulares, por lo que el objetivo es medir qué tanto los datos apoyan un escenario sobre otro. Los métodos de máxima verosimilitud pueden usarse para escoger un modelo o parámetros de un modelo que sean los que

apoyen mejor los datos. Antes, estos métodos asumían una forma particular de la distribución de las frecuencias alélicas entre poblaciones, pero más recientemente se han calculado con base en modelos de coalescencia. Beerli y Felsenstein (1999, 2001) desarrollaron un método que estima simultáneamente los tamaños efectivos de muchas poblaciones, así como las tasas de migración entre cada par de poblaciones. Este método puede implementarse con el programa MIGRATE. Hay otros métodos para inferir cambios históricos en la estructura poblacional que serán más apropiados que F_{st}, para estimar el flujo génico entre poblaciones grandes que actualmente experimentan niveles bajos de flujo génico.

Bibliografía

Avise, J.C. 1994. *Molecular markers, natural history and evolution*. Chapman and Hall, NY.

———. 2000. *Phylogeography: The History and Formation of Species*. Harvard University Press, Cambridge Ma.

Beerli, P. y J. Felsenstein. 1999. Maximum-likelihood estimation of migration rates and effective population numbers in two populations using a coalescent approach. *Genetics* 152**:**763-773.

———. 2001. Maximum likelihood estimation of a migration matrix and effective population sizes in *n* subpopulations by using a coalescent approach. *Proceedings of the National Academy of Sciences USA* 98:4563-4568.

Clement, M., D. Posada, K. Crandall. 2000. TCS: a computer program to estimate gene genealogies. *Molecular Ecology* 9:1657-1660.

Crow, J.F. y K. Aoki. 1984. Group selection for a polygenic behavioural trait: estimating the degree of population subdivision. *Proceedings of the National Academy of Sciences USA* 81:607.

Devlin, B. y N.C. Ellstrand. 1990. The development and application of a refined method for estimating gene flow from angiosperm paternity analysis. *Evolution* 44:248-259.

Gaggiotti, O.E., O. Lange, K. Rassmann, C. Gliddon. 1999. A comparison of two indirect methods for estimating average levels of gene flow using microsatellite data. *Molecular Ecology* 7:1513-1520.

Hedrick, P.W. 1999. Perspective: Highly variable loci and their interpretation in evolution and conservation. *Evolution* 53:313-318.

Levin, D.A. 1981. Dispersal versus gene flow in plants. *Annals of Missouri Botanical Garden* 68:233-253.

Mantel, N. 1967. The detection of disease clustering and generalized regression approach. *Cancer Research* 27:209-220.

Meagher, T.R. 1986. Analysis of paternity within a natural population of *Chamaelirium luteum.* 1. Identification of most-likely male parents. *American Naturalist* 128:199-215.

Nei, M. 1973. Analysis of gene diversity in subdivided populations. *Proceedings of the National Academy of Sciences* 70:3321-3323.

Neigel, J.E. 2002. Is F_{ST} obsolete? *Conservation Genetics* 3:167-173.

Posada, D., K.A. Crandall, A.R. Templeton. 2000. GEODIS: a program for the cladistic nested analysis of the geographical distribution of genetic haplotypes. *Mol. Ecol.* 9:487-488.

Riesberg, L.H. y J.M. Burke. 2001. The biological reality of species: gene flow, selection and collective evolution. *Taxon* 50:235-255.

Rosenberg, N.A. y M. Nordborg. 2002. Genealogical trees, coalescent theory and the analysis of genetic polymorphisms. *Nature Reviews Genetics* 3:380-390.

Roeder, K., B. Devlin, B.G. Lindsay. 1989. Application of maximum likelihood methods to population genetic data for the estimation of individual fertilities. *Biometrics* 45:363-379.

Schaal, B.A., D.A. Hayworth, K.M. Olsen, J.T. Rausher, W.A. Smith. 1998. Phylogeographic studies in plants: problems and prospects. *Molecular Ecology* 7:465-474.

Schaal, B.A., J.F. Gaskin, A.L. Caicedo. 2003. The Wilhelmine E. Key 2002 Invitational Lecture. Phylogeography, haplotype trees and invasive plant species. *Journal of Heredity* 94:197-204.

Slatkin, M. 1981. Estimating levels of gene flow in natural populations. *Genetics* 99:323-335.

———. 1985a. Gene flow in natural populations. *Annual Review of Ecology and Systematics* 16:393-430.

———. 1985b. Rare alleles as indicators of gene flow. *Evolution* 39:53-65.

———. 1991. Inbreeding coefficients and coalescence times. *Genetic Research* 58:167-165.

———. 1993. Isolation by distance in equilibrium and non-equilibrium populations. *Evolution* 47:264-279.

———. 1994. Gene flow and population structure. En: *Ecological Genetics* (ed Real LA). Pp. 3-18. Princeton University Press.

Slatkin, M. y N.H. Barton. 1989. A comparison of three indirect methods for estimating average levels of gene flow. *Evolution* 43:1349-1368.

Smouse, P.E y T.R. Meagher. 1994. Genetic analysis of male reproductive contributions in *Chamaelirium luteum* (L.) Gray (Liliaceae). *Genetics* 136:313-322.

Templeton, A.R. 1998. Nested clade analyses of phylogeographic data: testing hypotheses about gene flow and population history. *Molecular Ecology* 7:381-397.

———. 2004. Statistical phylogeography: methods of evaluating and minimizing inference errors. *Molecular Ecology* 13:789-809.

Templeton, A.R., E. Boerwinkle, C.F. Sing. 1987. A cladistic analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping. I. Basic theory and an analysis of alcohol dehydrogenase activity in Drosophila. *Genetics* 117:343-351.

Templeton, A.R., E. Routman, C. Phillips. 1995. Separating population structure from population history: a cladistic analysis of the geographical distribution of mitochondrial DNA haplotypes in the Tiger salamander, *Ambystoma tigrinum. Genetics* 140:767-782.

Whitlock, M.C. y D.E. McCauley. 1999. Indirect measures of gene flow and migration: FST not equal to 1/(4Nm + 1). *Heredity* 82:117-25.

Weir, B.S. y C.C. Cockerham. 1984. Estimating F-statistics for the analysis of population structure. *Evolution* 38:1358-1370.

Wright, S. 1943. Isolation by distance. *Genetics* 28:114-138.

———. 1951. The genetical structure of populations. *Annals of Eugenics* 15:323-354.

———. 1969. *Evolution and Genetics of Populations*, vol.2, *The theory of gene frequencies.* University of Chicago Press, Chicago.

Capítulo 3

Tamaño efectivo de la población

Alejandra Moreno Letelier

El tamaño efectivo poblacional fue mencionado por primera vez por S. Wright en su trabajo sobre deriva génica, en el que considera a éste como el factor determinante de la tasa de decaimiento de la heterocigosis: el número de heterocigotos en una población disminuirá a una tasa de (½N), donde N es el tamaño poblacional (Wright, 1930).

En el proceso de deriva génica, los alelos presentes en un momento dado en una población de individuos diploides son sólo una muestra de los alelos presentes en la generación previa, ya que algunos de ellos no fueron transmitidos a la siguiente generación únicamente por azar, y así sucesivamente hacia el pasado (Futuyma, 1986). Con el paso del tiempo, las siguientes generaciones presentarían cada vez menos copias de los alelos originales y cada vez los individuos se volverían más homócigos, aumentando también la probabilidad de que sus alelos sean idénticos por descendencia (ver capítulo 6). En este proceso gobernado por el azar, la pérdida de heterocigocis va también acompañada de cambios en las frecuencias alélicas, ya que un alelo que por azar deje más descendencia aumentaría su frecuencia en la población.

Puede entonces considerarse que la probabilidad de que un alelo esté representado en la siguiente generación por i número de copias, depende del número de copias total (2N, donde N es el tamaño de la población) de la generación parental y las frecuencias de los alelos. Eventualmente el muestreo azaroso de alelos de una generación a otra llevará a la fijación de una de las

copias (frecuencia de 1) y a la pérdida de la otra (frecuencia de 0) generando conjuntamente la pérdida de heterocigocis en la población (Futuyma, 1986).

La tasa de pérdida de variación y de heterocigosis por deriva génica es más alta en poblaciones pequeñas, aunque ocurre en todas las poblaciones finitas, a menos que la mutación y la inmigración contrarresten el proceso y lo hagan más lento. Por ello resulta necesario tener un estimado confiable del tamaño de la población para entender el proceso de la deriva génica y la pérdida de variación.

¿Qué significa el tamaño efectivo?

La base intuitiva del concepto de tamaño efectivo de la población (N_e) propuesto por Sewall Wright está asociada al tamaño poblacional que es relevante en términos evolutivos, esto es, el número de individuos reproductivos, ya que son éstos los que contribuyen a la generación siguiente en términos demográficos y sobre todo genéticos (Wright, 1969; Hedrick, 2000). Sin embargo, no existe una definición muy precisa de este concepto, por lo que la mayoría de las definiciones se refiere al número de individuos en una población cuya heterocigosis decrece a una tasa $k=1/2N$, donde N es el número poblacional (Merrel, 1981). Dicha población, considerada como ideal, debe cumplir las siguientes condiciones:

- El número de hembras debe ser igual al de machos.
- El apareamiento es al azar (panmixia).
- Todos los individuos contribuyen genéticamente a la siguiente generación.
- La fecundidad por familia tiene una distribución de Poisson, es decir, que los individuos reproductivos únicamente se reemplazan en la población a la siguiente generación.
- La población está en equilibrio de Hardy-Weinberg.

Si una población cumple con *todos* estos requisitos, el tamaño efectivo es igual al tamaño censal. Sin embargo las variaciones en uno de estos supuestos (o en varios) afectan el tamaño efectivo de la población y por consiguiente, la forma de calcularlo, como veremos a continuación.

Métodos de estimación directos

Sexos separados

Si se considera un organismo dioico, la mitad de los gametos proviene de los individuos de un solo sexo, por lo que el tamaño efectivo depende del número de individuos de cada uno:

$1/N_e = ¼\ N_h + ¼\ N_m$ lo que significa $N_e = 4N_hN_m/\ N_h + N_m$

donde el número restrictivo es el del sexo menos numeroso (Merrel, 1981; Hedrick, 2000).

Fluctuaciones temporales en el tamaño poblacional

La mayoría de los estimados de N_e se basa en que la población permanece estable en el tiempo, sin embargo, cuando existe variación cíclica a través del tiempo que afecte los tamaños demográficos, el tamaño efectivo de la población a largo plazo también se verá afectado por los tamaños poblacionales menores, ya que los descendientes de un periodo de tamaño poblacional muy pequeño, serán los sobrevivientes de un cuello de botella. En este caso de poblaciones fluctuantes Sewall Wright considera la media armónica de los tamaños poblacionales como el tamaño efectivo de la población:

$$1/N_e = 1/T\ [1/N_1 + 1/N_2 + 1/N_3 + \ldots\ 1/N_T]$$

donde T es el número de generaciones (Merrel, 1981).

Endogamia

La endogamia se refiere al apareamiento no aleatorio en donde la probabilidad de que un individuo sea autócigo (con alelos idénticos por descendencia) es mayor que para una población panmíctica. Por lo tanto, la endogamia representa una violación al supuesto de la panmixia, y reduce el número real de individuos reproductivos. Además, una endogamia intensa, producida por ejemplo por autofertilización, también reduce la heterocigosis, y el tamaño poblacional es más parecido al de una población haploide (Hedrick, 2000).

En este caso: $N_e = N/(1+ F)$, donde F es el coeficiente de endogamia (véase el capítulo 6 de este libro).

Otros métodos de estimación

Un problema de los primeros índices del tamaño efectivo fue que muchos de los parámetros eran difíciles de estimar en campo, además de que no consideraban la migración, las diferencias en historias de vida de las especies o la estructuración poblacional (ver capítulo 2). A partir de esos primeros índices se han desarrollado muchos otros, con grados variables de complejidad, que pueden lidiar con una o varias violaciones al modelo de población ideal. Es muy importante considerar que la eficacia y la exactitud de estos métodos dependen de la información disponible para las poblaciones y de la historia natural particular de la especie que se esté trabajando.

Métodos ecológicos y demográficos

Métodos para poblaciones aisladas y generaciones sobrelapadas

Estructura de edades

En especies que tienen poblaciones de individuos prereproductivos y postreproductivos, la tasa de nacimientos y muertes puede variar por factores ambientales y genéticos dependiendo del grupo de edad al que pertenezcan. Entonces, los individuos de distintas edades tendrán una contribución diferencial a las siguientes generaciones (Nei e Imazumi, 1966). Estas consideraciones demográficas pueden ser integradas a métodos para calcular N_e:

$$N_e = N_r L$$

donde N_r es el número de individuos nacidos en un año dado que llegan a reproducirse y L es el tiempo generacional (Eguiarte *et al.*, 1993; Hedrick, 2000).

Estructura de edades y diferencias en fecundidad

Este método es uno de los más clásicos y considera el número de individuos nacidos en un año dado (N_o), el tiempo generacional (L) y el parámetro i que

significa la suma de todas las edades del cuadrado de la sobrevivencia específica de cada edad por la fecundidad específica para las edades:

$$N_e = N_o L_i$$

Se supone una estructura estable de edades y que la población no está creciendo (Crow y Kimura, 1984; Eguiarte *et al.*, 1993).

Diferencias en el éxito reproductivo

Este método es uno de los más empleados y confiables en el cálculo de N_e, ya que aborda una de las violaciones más comunes de la población ideal: que la progenie por familia no tenga una distribución de Poisson (Eguiarte *et al.*, 1993; Hedrick, 2000). El supuesto principal es que la población tenga una estructura estable de edades y un tamaño constante. El tamaño efectivo se calcula:

$$Ne = \frac{4(N_0 - 2)\, L}{V + 2}$$

donde N_o es el número de individuos nacidos en una temporada, L el tiempo generacional y V es la varianza en el tamaño de las familias (Eguiarte *et al.*, 1993). Sin embargo, muchos de estos parámetros son difíciles de obtener para especies con generaciones muy largas (Storz *et al.*, 2001).

Existe una variante que considera la varianza en el éxito reproductivo específico de cada sexo, lo que permite el estudio de sus efectos sobre el tamaño efectivo para cada sexo y para loci heredados uniparentalmente (mitocondria, cloroplasto o cromosoma Y). Por ejemplo, en especies que presentan poliginia la varianza del éxito reproductivo de los machos es mucho mayor que la de las hembras, por lo que su tamaño efectivo es menor. Lo mismo para loci heredados uniparentalmente, cuyo tamaño efectivo es 4 veces menor al de los loci autosomales (Storz *et al.*, 2001). El cálculo del tamaño efectivo para genes autosomales es:

$$\frac{1}{N_e A} = \frac{1}{16N_m}\left[\, 2 + var\,(X_{mm}) + 2\left(\frac{N_m}{N_f}\right) cov(X_{mm,mf}) + \left(\frac{N_m}{N_f}\right)^2 var\,(X_{mf})\right] + \frac{1}{16N_f}\left[\, 2 + var\,(X_{ff}) + 2\left(\frac{N_f}{N_m}\right) cov(X_{fm,ff}) + \left(\frac{N_f}{N_m}\right) 2\, var\,(X_{fm})\right]$$

donde N_i es el número de individuos de sexo *i* (*i*=masculino o femenino) y X_{ij} es la progenie neta (de toda la vida) de sexo *j* (*j*=masculino o femenino) para cada padre de sexo *i*. Para loci mitocondriales y de cromosoma Y las fórmulas son:

$$\frac{1}{N_e Y} = \frac{\frac{\{2+2[var(X_{mm})]\}}{N_m}}{\frac{\{2+2[var(X_{ff})]\}}{N_f}}$$

En este método, los datos individuales entran en una simulación de cerca de 1000 repeticiones para obtener los estimados de N_e.

En el caso de que la población esté creciendo puede usarse un método basado en la varianza del éxito reproductivo que es :

$$N_e = \frac{Nk - 1}{(k - 1) + \frac{V_k}{k}}$$

donde *k* es el número promedio de la progenie por cada padre y V_k es la varianza en el éxito reproductivo (Hedrick, 2000).

Varianza del éxito reproductivo

Este método requiere de estimados del potencial reproductivo parcial (de una temporada) y la media y varianza de la sobrevivencia de los individuos. A partir de estos estimados, el tamaño efectivo poblacional basado en el potencial reproductivo a lo largo de la vida N_{lrs}, se calcula:

$$N_{lrs} = \frac{4N - 4}{S^2_{vlrs} + 2}$$

donde *N* es la población muestreada y S^2_{vlrs} es la varianza en el éxito reproductivo a lo largo de la vida. En poblaciones pequeñas, el que algunos individuos vivan un poco más aumenta muchísimo la varianza en el éxito reproductivo (Barrowclough y Rockwell, 1993; Tremblay y Ackerman, 2001).

Cuando el cálculo del éxito reproductivo a lo largo de la vida se complica por tiempos generacionales muy largos, se pueden aplicar simulaciones basadas en individuos que además permiten evaluar los efectos de la estocasticidad demográfica. En este caso se necesita información demográfica como las tasas de sobrevivencia y fertilidad específicas de cada edad para simular el destino reproductivo de cada individuo en una cohorte (Stroz *et al.*, 2001). También pueden hacerse simulaciones del éxito reproductivo y sobrevivencia para cada sexo. Con base en estas simulaciones se obtiene el éxito reproductivo para cada individuo en su tiempo de vida, lo cual es directamente aplicable al método de Barrowclough y Rockwell (Tremblay y Ackerman, 2001).

Método de Orive

Utiliza las variables demográficas descritas anteriormente pero además incluye un modelo de matrices de estructura de edades o tamaños. El estimado del tamaño efectivo se basa en el tiempo de coalescencia y es muy útil para organismos diploides, con reproducción sexual y generaciones sobrelapadas. Sus supuestos son que el tamaño poblacional debe ser constante y con una estructura de edades estable. El estimado de tamaño efectivo depende de la probabilidad de reproducción y de transición entre las categorías para cada población (Orive, 1993; Tremblay y Ackerman, 2001).

Método de Nunney

Calcula el cociente del tamaño efectivo entre el tamaño censal (N_e/N) de una población y utiliza datos demográficos. Es necesario que la especie en cuestión cumpla con los supuestos de apareamiento aleatorio (panmixia), que la sobrevivencia de machos y hembras sea independiente después de la madurez sexual, que el éxito reproductivo tenga una distribución de Poisson y que la proporción de adultos que intenta reproducirse sea conocida.

$$\frac{N_e}{N} = \frac{4r(1\text{-}r)T}{rA_f(1+I_{Af}) + (1\text{-}r)\,A_m\,(1 + A_m) + (1 - r = I_{bm} + rI_{bf}}$$

Los parámetros son los siguientes: tiempo de maduración (M), longevidad de los adultos (A), tiempo generacional (T), varianza en el éxito reproductivo de las hembras (I_{bf}), varianza en el éxito reproductivo de los machos (I_{bm}) y

proporciones sexuales de los adultos (*r*). Un aspecto que puede influenciar los cálculos es el sistema de apareamiento. En general, para ciertos animales se asume que los machos se aparean varias veces con un número aleatorio de hembras, y que las hembras lo hacen una sola vez. Apareamientos múltiples por parte de las hembras aumentaría los estimados de N_e (Antolin *et al.*, 2001; Nunney, 2002).

Métodos para poblaciones aisladas y generaciones discretas

El método de Heywood (1986) se basa en el de Kimura y Crow (1970) que describe el tamaño efectivo como una función de la distribución de la progenie entre los padres. El modelo de Heywood reformula el anterior, considerando la varianza en el éxito reproductivo. Este método es adecuado para especies con generaciones discretas, donde la varianza del éxito reproductivo asociada al tiempo de vida es mínima, ya que todos los individuos tienen periodos de vida cortos y muy similares. La estimación se hace de la siguiente manera:

$$Nv = \left[\frac{1}{(1 + F)\,(s_b^2/X_b^2) + 1}\right] N$$

donde *b* es la variable con la cual se mide el éxito reproductivo (número de flores, número de semillas, número de hijos, etc.), s^2 y $\bar{X}$ son la varianza y la media de la variable, *F* es el coeficiente de endogamia y *N* el número de muestra (Tremblay y Ackerman, 2001).

Otro método aplicable a especies con generaciones discretas es el de Waples, especialmente para especies que presentan bancos de semillas o tiempos de maduración sexual variables, ya que esto actúa como una forma de almacenamiento genético. Incluye también la posibilidad de que el tamaño de la población sea fluctuante y se expresa como $N_e = TN_h$, donde *T* es el tiempo promedio a la madurez sexual (Nunney, 2002).

Método para poblaciones continuas: tamaño del vecindario

En muchos casos, no se pueden aplicar los métodos anteriores, ya que no es posible delimitar claramente las poblaciones que tienen una distribución muy grande y continua. Para salvar este problema se introdujo el concepto de tamaño del vecindario (Wright, 1943) que permite evaluar los efectos de

tamaños finitos en especies con este tipo de distribución (Hedrick, 2000). El tamaño del vecindario está determinado por la capacidad de dispersión del organismo en cuestión, que en el caso de animales puede ser medida directamente (aunque en animales migratorios puede complicarse). En plantas tiene que considerarse la tasa de fecundación cruzada, la dispersión del polen y la de las semillas (Eguiarte *et al.*, 1993; véanse los capítulo 5 y 6 de este libro). El tamaño de vecindario es el número de individuos en un círculo cuyo radio es el doble de la desviación estándar del flujo génico por generación ($\sqrt{2V}$). Si se conoce la densidad de los individuos (d) y el área del círculo del vecindario ($4\pi V$), se puede estimar el tamaño efectivo de la población como sigue (Hedrick, 2000):

$$Ne = 2\ \pi\ Vd$$

A veces, el cálculo del tamaño efectivo utilizando este método no es lo suficientemente preciso debido a otros factores que desvían a la población del ideal. Sin embargo, conociendo el tamaño del vecindario, es posible considerarlo como una población aislada y aplicar cualquiera de los otros métodos desarrollados para este tipo de poblaciones, pero cuyos supuestos se apeguen mejor a las características de la especie bajo estudio (Eguiarte *et al.*, 1993).

Métodos indirectos

Como vimos anteriormente, se pueden obtener estimados de N_e con métodos ecológicos utilizando información demográfica. Estos estimados nos dan como resultado el tamaño efectivo actual, en tiempo ecológico (instantáneo) y muchas veces, de una generación a otra. Por su parte, los estimados de tamaño efectivo con métodos genéticos son un promedio a lo largo de un periodo de tiempo que está determinado por la tasa de mutación por generación del locus bajo estudio (Beerli y Felsentein, 1999).

Otros métodos pueden estimar los tamaños efectivos a corto plazo al analizar los efectos de la deriva génica sobre las medidas genéticas, ya que si el tamaño efectivo histórico de la población ha sido pequeño, se espera encontrar más evidencia de deriva génica, mientras que si el tamaño efectivo ha sido grande no se obtendrá mucha señal de deriva en los datos (Hedrick, 2000). Pueden entonces compararse con diversos métodos los datos reales con lo esperado por deriva génica para estimar el tamaño efectivo de la población. A continuación enlistamos los más utilizados:

Métodos de medición de cambios genéticos

Como se mencionó anteriormente, es posible inferir N_e a partir de los cambios en la composición genética de una población a través del tiempo. Por ejemplo la disminución de la heterocigocis debida a la deriva génica sería:

$$H_e(L) = H_e(0)\left(1 - \frac{1}{2N_e}\right)^L$$

o bien, con base en las fluctuaciones temporales de las frecuencias génicas, donde $P(L)$ es la frecuencia génica al tiempo L (con $Q = 1\text{-}P$) :

$$\frac{1}{2N_e} = \frac{[P(L) - P(0)]^2}{LP(0)Q(0)}$$

Estimados basados en F_{ST}

La estructuración poblacional es uno de los factores que pueden influenciar N_e. El modelo más sencillo para ejemplificar esto es el modelo de islas, donde la población está subdividida en un número aleatorio de islas con reproducción panmíctica dentro de ellas, salvo por la recepción ocasional de migrantes, todos ellos con la misma probabilidad de llegar a una de las islas desde cualquier otra (figura 1; Wright, 1969; ver capítulo 2). En este caso, se espera que la subdivisión aumente N_e en toda la población (Laporte y Charlesworth, 2002), como lo muestra la fórmula: $N_e=ND/(1-F_{st})$, donde N es el número censal y D el número de subpoblaciones. En cualquier caso F_{st} puede calcularse independientemente usando datos de fruecuencias génicas (Nichols *et al.,* 2001):

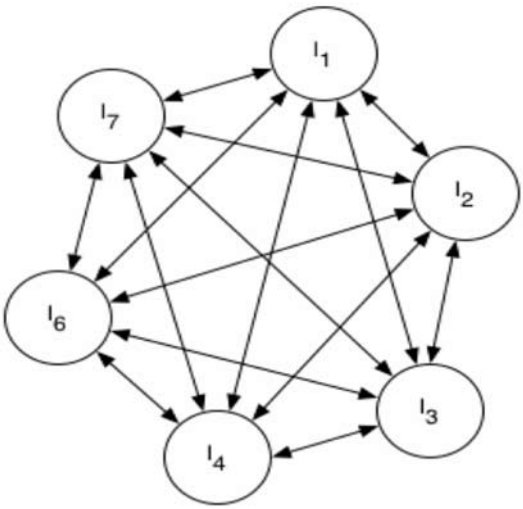
$$F_{ST} = \frac{2Var(p_i)}{1 - G_T}$$

donde $Var(p_i)$ es la varianza del alelo p en i número de subpoblaciones y G_T es la probabilidad de que dos alelos tomados al azar (con reemplazamiento) de la especie completa sean idénticos por descendencia: $G_T = p^2 + q^2$ (Gillespie, 1998).

Sin embargo, la subdivisión poblacional puede reducir el tamaño efectivo poblacional cuando cada una de las islas contribuye de manera desigual a la poza de migrantes. Considerando estos casos se han desarrollado otros métodos como

el modelo de *stepping stone* o piedras de paso. En este modelo $N_b=2\pi Nm$, donde N_b es el tamaño efectivo poblacional del vecindario (Slatkin y Barton, 1989).

Figura 1. El modelo de islas de Wright asume tamaño poblacional igual y que el intercambio de migrantes de una isla a otra ocurre con la misma pobabilidad (*m*)



Métodos de máxima verosimilitud y coalescencia

La máxima verosimilitud de un parámetro discreto dado un conjunto de datos, es la probabilidad de observar esos datos dado ese valor del parámetro (Williamson y Slatkin, 1999; ver capítulo 4).

Este tipo de métodos se ha desarrollado mucho en los últimos años, sobre todo porque ofrecen estimados más robustos que los métodos tradicionales basados en F_{st}, además de que algunos de estos métodos pueden incluir la estimación de eventos de migración, cambios de tamaño poblacional y otras violaciones al modelo de población ideal (Berthier *et al.*, 2002; Beerli y Felsenstein, 1999; Kuhner *et al.*, 1998; Williamson y Slatkin, 1999).

Uno de los modelos más sencillos es el que considera las frecuencias alélicas de loci neutrales y no ligados tomados al azar de una población, a generaciones t_0, t_1, t_2,...t_f. Esto se basa en el modelo de Wright-Fisher en que la varianza

en las frecuencias alélicas con valor inicial de p es $p(1-p)/2N_e$ después de una generación, por lo tanto, si se conoce la varianza de las frecuencias alélicas en distintas generaciones, es posible aproximar N_e (Williamson y Slatkin, 1999). Este método, como la mayoría de los métodos basados en los estadísticos F, tiende a sobreestimar N_e, pero posee intervalos de confianza menores y posee la flexibilidad suficiente para modificar alguno de los supuestos, como el tamaño poblacional, ya que permite incluir datos de crecimiento exponencial de la población. Sin embargo, es muy demandante computacionalmente y no puede lidiar con eventos de selección, migración y mutación que también pudieran cambiar las frecuencias alélicas (Williamson y Slatkin, 1999).

Como alternativa a los métodos de máxima verosimilitud tradicionales, existen los métodos que utilizan el algoritmo de aproximación Metropolis-Hastings cadena de Markov Monte-Carlo y que generan todas las posibles genealogías y eventos de migración para buscar el escenario que mejor explique los datos, además de que acepta distintos modelos de mutación: modelo de alelos infinitos para marcadores electroforéticos; el modelo paso a paso para microsatélites y el modelo de sitios finitos para secuencias de ADN (Beerli y Felsenstein, 1999). Este modelo consiste en dos poblaciones con tamaños efectivos $N_{e(1)}$ y $N_{e(2)}$ (que no tienen que ser iguales), y que intercambian migrantes a tasas m_1 y m_2. Los parámetros que se calculan son:

$$\bar{\gamma} = 4N_e m$$

donde $\bar{\gamma}$ es el número efectivo de migrantes, N_e el tamaño efectivo de la población y m es la tasa de migración por generación.

Adicionalmente se calculan otros parámetros que "absorben" el de tasa de mutación que generalmente no se conoce como $\theta = 4N_{e(i)}m$ y $M_i = m_i/m$ (Beerli y Felsenstein, 1999), que puede luego sustituirse en:

$$\gamma = \theta_i M_i = 4N_{e(i)} m_i$$

A partir de esto, se aplican métodos de máxima verosimilitud.

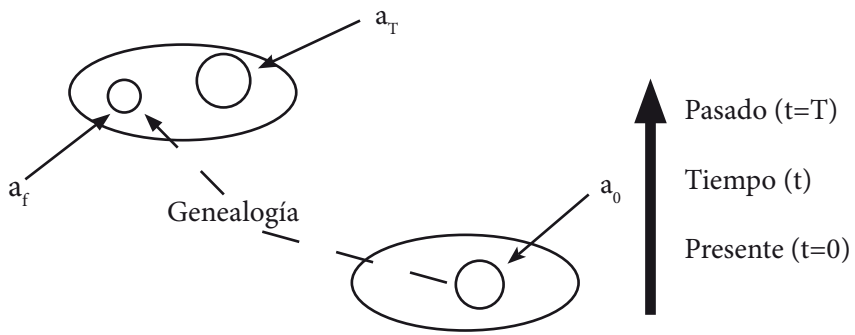
En métodos anteriores no se podían obtener estimados de coalescencia en presencia de migración, pero el presente método especifica los momentos y lugares de los eventos de migración. Cabe notar aquí que los parámetros iniciales necesarios para calcular N_e y los eventos de migración (θ y M), son obtenidos de manera independiente usando F_{st}. Este método inicialmente no admitía variaciones temporales en el tamaño poblacional, pero una variante del

mismo con un muestreo de genealogías Metropolis-Hastings permite calcular θ y la tasa de crecimiento de la población (Kuhner *et al.*, 1998).

Este conjunto de métodos son utilizados por el programa LAMARC (http://evolution.genetics.washington.edu/lamarc.html/) y sus distintas variantes dependiendo de los supuestos iniciales y de los estimados de interés.

Existen otros métodos que utilizan máxima verosimilitud y coalescencia, como el propuesto por Berthier *et al.* (2002), en el que se toman dos muestras de una población cerrada en distintos tiempos (figura 2).

Figura 2. Modelo de máxima verosimilitud . La población es la línea punteada, a_0 es la muestra más reciente (t=0) y la muestra más antigua es aT tomada al tiempo t=T. Se supone que a_0 tiene una genealogía y a_f representa sus fundadores



Para un locus, cada una de las muestras representa un vector (a_0 y a_T), de n_0 y n_T cromosomas con k alelos distintos. Para el número de alelos distintos se considera el total de los observados en las muestras, donde cada una de las muestras tiene <k alelos (k_0 y k_T). Se supone que las muestras son independientes con probabilidades $p(a_T|n_T, k, x)$ y $p(a_0|t, N_e, n_0, k, x)$, donde N_e es el tamaño efectivo de la población y x es la frecuencia genética paramétrica al tiempo T. Las muestras en cada loci m también se suponen independientes y por lo tanto, su verosimilitud general sería L (T, N_e, x_1, x_2, ..., x_m), que se obtiene multiplicando las probabilidades a través de todas las muestras y a través de todos los loci (Berthier *et al.*, 2002). También se utiliza el algoritmo Metropolis-Hastings para integrar x_i. Cabe hacer notar que este método no funciona en presencia de migración, pero sí tiene mayor precisión que los métodos basados en F_{st} (Berthier *et al.*, 2002).

También se han hecho algunas modificaciones a los métodos de máxima verosimilitud, siendo una de ellas el llamado método de pseudo-máxima verosimilitud (Seo *et al.*, 2002). Es llamado así porque en algunos métodos comunes de máxima verosimilitud, los intervalos de tiempo a la coalescencia son estimados y no conocidos con certeza, lo que añade ruido a los cálculos. El método de pseudo-máxima verosimilitud calcula *t* independientemente y luego es incluido a la función de máxima verosimilitud.

$$\hat{N}_e = \frac{1}{2(n-1)} \sum_{i=2}^{2} i\,(i-1)t_i$$

siendo *n* el número de secuencias, *i* el número de linajes ancestrales y t_i la duración del periodo de tiempo con *i* linajes.

Sin embargo, este método no puede tomar en cuenta los niveles de incertidumbre de los parámetros calculados independientemente, como por ejemplo los de la genealogía, para hacerlo sería necesario adoptar una aproximación bayesiana (Seo *et al.*, 2002; Drummond *et al.*, 2002).

Métodos de coalescencia

A partir de datos genéticos es posible calcular el tiempo de divergencia entre alelos dentro de una población usando como base la teoría de coalescencia. Con estos análisis se puede determinar la genealogía de los alelos, el tiempo hacia el pasado en que divergieron de uno solo y parámetros demográficos como el tamaño efectivo histórico de la población (Furlong y Brookfield, 2001). Esto es posible debido a que la probabilidad de coalescencia varía de forma recíproca al tamaño efectivo poblacional, como lo muestra la fórmula $\tau = N_e\,(1-1/n)$ donde τ es el número de generaciones a la coalescencia de *n* alelos de un locus muestreado al azar en una población (Hughes y Verra, 2001). También puede calcularse el tiempo de coalescencia de manera independiente con base en las tasas de sustitución de nucleótidos de una secuencia: $l = \pi_s/2\lambda_s$, donde π_s es el número de sustituciones sinónimas pareadas y λ_s es la tasa de sustitución (Hughes y Verra, 2001).

Los tiempos entre eventos de coalescencia están determinados por distribuciones exponenciales negativas independientes, donde para un evento de coalescencia en que *i* linajes disminuyen a *i-1*, entonces $t = 2N[i(i-1)]$, donde

N es el tamaño efectivo haploide de la población. En este caso *N* debe ser constante a través del tiempo.

Sin embargo, estos primeros análisis son muy limitados en cuanto a los supuestos de tamaño poblacional constante y a la ausencia de selección natural. Violaciones a este punto pueden tener distintas implicaciones: por ejemplo en el caso de un barrido selectivo o *selective sweep* todos los alelos serían producto de un solo alelo ventajoso de origen muy reciente, lo que generaría un patrón similar a la expansión poblacional (Furlong y Brookfield, 2001). Aunque también puede darse el caso de que, en efecto, haya existido una expansión demográfica en el pasado, y al contrario de una población constante, los eventos de coalescencia no son independientes entre sí, sino que unos proceden de otros en un periodo de tiempo muy corto, dando como resultado tiempos de coalescencia pequeños y ausencia de linajes que divergieron muy temprano en el tiempo (Furlong y Brookfield, 2001).

Un ejemplo de cómo se puede calcular el tiempo de coalescencia relacionado con N_e en casos de que la población haya crecido en el pasado fue desarrollado originalmente por Slatkin y Hudson:

$$t = \frac{ln}{r}1 - \left[\frac{(2rN\, ln(1 - P))}{n(n - 1)}\right]$$

donde *P* es la probabilidad de que la coalescencia pase en *t* generaciones, *r* la tasa de crecimiento poblacional y *n* el tamaño muestral (Furlong y Brookfield, 2001).

Existen también otros métodos basados en coalescencia que pueden ser aplicados a poblaciones con otras violaciones a la población ideal, por ejemplo el desarrollado por Laporte y Charlesworth (2002) para poblaciones subestructuradas y para loci con distintos patrones de herencia. Bajo este modelo se pueden estimar los tiempos de coalescencia pareados y probabilidades de identidad por descendencia de alelos en una metapoblación con un esquema de migración arbitrario. Se basa en que la probabilidad de que un par de genes coalesca es de: $P_i = 1/2t_iN_{ie}$.

Los individuos en cada deme (grupo de individuos con reproducción panmíctica) son divididos en machos y hembras, que tienen diferentes propiedades demográficas y migratorias. A partir de lo anterior se pueden obtener estimados de N_e con fuerte base biológica. Por ejemplo, para loci autosomales, la expresión sería:

$$\frac{1}{N_{eA}} \approx \frac{(1+F)}{4}\left\{\frac{1}{N_f}+\frac{1}{N_m}+\frac{(1-c)^2\,\Delta V_f}{N_f}+\frac{c^2\,\Delta V_f}{N_m}\right\}$$

y para genes con herencia materna:

$$\frac{1}{N_{eA}} \approx \frac{2\,[1+(1-c)^2\,\Delta V_f]}{N_f}$$

donde ΔV_f y ΔV_m son los excesos sobre una distribución de Poisson del total de progenie *per capita* para hembras y machos respectivamente; F es el coeficiente de endogamia; N_f y N_m el número de hembras y machos; y c es la proporción de machos en los adultos reproductivos (Laporte y Charlesworth, 2002). Cabe hacer notar que este método, aunque basado en coalescencia, requiere ciertos datos demográficos para estimar las probabilidades de coalescencia y las desviaciones a la distribución de Poisson en la progenie. Sin embargo, es bastante versátil porque permite el cálculo de la coalescencia y de tamaño efectivo para distintos escenarios y con diferentes tipos de datos, además de que es más adecuado que los métodos clásicos de F_{st} para estimar tamaños efectivos en presencia de migración específica a un sexo (p.ej. filopatría) y de fuerte diferenciación entre subpoblaciones (Laporte y Charlesworth, 2002).

Genes con herencia uniparental

Genes como los de mitocondria, cloroplasto y cromosoma Y han sido muy utilizados en años recientes para establecer relaciones entre organismos y para estudios de genética de poblaciones (Hedrick, 2000). Han resultado especialmente útiles en disciplinas como la filogeografía, ya que al ser haploides permiten reconstruir filogenias sin el inconveniente de la recombinación (que genera reticulación de ramas) y estimar tiempos de coalescencia con mayor exactitud (ver capítulo 15).

En el caso de estos genes, si la proporción de los sexos es igual, su tamaño efectivo es igual a $N_e/4$, ya que sólo se heredan por uno de los padres y cada uno de los padres sólo tiene una versión del gen. Sin embargo, si la proporción de sexos no es igual, en el caso de genes de herencia materna (mitocondria y cloroplasto en la mayoría de los casos) el tamaño efectivo es: $N_e = N_{ef}/2$, que

significa la mitad del tamaño efectivo para las hembras. Si consideramos el cromosoma Y o el cloroplasto cuando se hereda paternalmente (como en coníferas) $N_e = N_{em}/2$ (Hedrick, 2000). Lo anterior sólo es válido si no existen diferencias ligadas al sexo en las tasas de mutación o si el éxito reproductivo o la tasa de migración es similar entre los sexos. Por ejemplo, si existe una varianza muy grande en el éxito reproductivo de los machos aumentará el tamaño efectivo para los genes del cromosoma X en comparación con los genes autosomales, ya que la transmisión del cromosoma X es el doble para hembras que para machos (Charlesworth, 2001).

Implicaciones evolutivas

Dado que la variación genética es la materia prima sobre la cual puede actuar la selección natural y dar origen a adaptaciones, se ha llegado a considerar su presencia como indicador de la buena "salud" de una población, mientras que su ausencia limitaría la capacidad de respuesta de una población a un medio cambiante en el corto y largo plazo (Amos y Harwood, 1998; véase el capítulo 8 de este libro). Como se mencionó al inicio de este capítulo, el tamaño efectivo de la población determina la tasa de pérdida de esta variación por deriva génica, fenómeno que puede calcularse es:

$$H_t = H_0 (1-1/2N_e)^t$$

donde H_0 y H_t son la heterocigosis inicial y al tiempo t, respectivamente.

En este contexto, es de esperarse que un gran tamaño efectivo reduzca la pérdida de variación por deriva génica y se mantenga la diversidad por largo tiempo, así como la efectividad de la selección natural sobre dicha variación (Cruzan, 2001; Kawata, 2001). Sin embargo los niveles de variabilidad no siempre corresponden a lo predicho por la teoría, ya que algunas especies con alta variación tienen poblaciones pequeñas y demográficamente en peligro, mientras que otras con crecimiento exponencial sólo presentan unos pocos genotipos distintos, como las plagas, lo que hace que la relación entre N_e y el nivel de pérdida de variabilidad sea más complejo (Amos y Harwood, 1998). Por ejemplo, durante los procesos de colonización y expansión del rango de distribución de una especie, pueden ocurrir muchos efectos fundador en las poblaciones pioneras, reduciéndose N_e a pesar de que el tamaño censal sea grande (Nichols *et al.*, 2001). Esto puede reducir la variación genética de toda la especie al terminar el proceso de expansión del rango a niveles menores que

lo esperado por el tamaño histórico de la especie en su conjunto. Lo mismo se observa en metapoblaciones, ya que en cada subpoblación N_e es generalmente muy pequeño y en el proceso de extinción y recolonización puede perderse mucha variación.

Otro efecto de la reducción de N_e y que cobra gran importancia a la luz de la biología de la conservación es la depresión por endogamia, que es la reducción en la adecuación de los individuos producto de apareamientos entre parientes, lo que expone los alelos deletéreos a la acción de la selección natural y puede reducir aún más la variación genética, más allá que la deriva génica por sí sola (Amos y Harwood, 1998; Thévenon y Couvet, 2002). Y si aunado a esto, la acción de la deriva génica es muy severa, estos alelos deletéreos pueden llegar a fijarse en la población y a la larga reducir la adecuación de la población entera y no de unos pocos individuos (Cruzan, 2001).

La variación genética se produce en primera instancia por mutación y bajo el modelo de mutación de sitios infinitos; el equilibrio de la variación neutral en una población panmíctica y cerrada está determinado por el producto de la tasa de mutación y el tamaño efectivo de la población (Cherry, 1998; Charlesworth, 2001). Con base en esto se han relacionado el tamaño efectivo y la tasa de sustitución (k) (Cherry, 1998; Seo *et al.*, 2002). Dentro del marco de la teoría casi neutral de la evolución, la probabilidad de fijación de un alelo mutante con un coeficiente de selección (s) depende del tamaño efectivo de la población (N_e) y es igual a: $S = 2N_e s$ para una población haploide. Los supuestos de esta afirmación son que existe una distribución constante de s entre mutantes y la tasa de mutación también es constante. Entonces, si se considera una distribución exponencial de s la relación es $k \propto 1/N_e$ y si se considera una distribución gamma, $k \propto 1/\sqrt{N_e}$ (Cherry, 1998). Esto predice que la probabilidad de fijación de las mutaciones favorables con coeficientes de selección pequeños va a ser mayor en poblaciones con N_e alto, y menor en poblaciones con N_e pequeño ya que en estos casos la deriva génica actúa con mayor fuerza que la selección, y dichas mutaciones se comportan como neutras. Lo mismo sucede con las mutaciones ligeramente deletéreas, que en poblaciones grandes son eliminadas por la selección, mientras que en poblaciones pequeñas se comportan como neutras (Cherry, 1998).

Si se considera un modelo donde la mayoría de las mutaciones son deletéreas y los coeficientes de selección tienen una distribución constante, en una población pequeña la deriva fijará una fracción de ellas, generando un decremento en la adecuación, mientras que en una población grande se eliminarán, generando un incremento en la adecuación (Cherry, 1998). Pero

esta afirmación es poco realista, ya que en algún momento debe alcanzarse un balance que mantenga la adecuación en un estado estable. Uno de los mecanismos para mantener ese balance considera la distribución de los coeficientes de selección una vez ocurrida la mutación, basándose en la idea de que si se han acumulado muchas mutaciones deletéreas en una secuencia codificante, quedan ya pocas maneras de empeorar la adecuación de esa secuencia y más formas de mejorarla (Cherry, 1998). Por otro lado, es necesario también considerar que aunque únicamente la mutación puede generar nueva variación dentro de una población, la migración puede aportar esa variación proveniente de otros sitios, por lo que no sólo el tamaño efectivo de la población es determinante para el mantenimiento de la variación genética y la adecuación, sino también la posibilidad de intercambiar genes por migración de individuos o gametos de poblaciones cercanas. Esto hace posible que poblaciones muy pequeñas puedan persistir a lo largo del tiempo gracias al efecto de rescate de los migrantes (Cruzan, 2001).

Consideraciones prácticas sobre los métodos indirectos

Los métodos indirectos se basan principalmente en el uso de marcadores moleculares para estimar la variación neutral a nivel intraespecífico, por lo que sus resultados dependen en gran medida de que estos marcadores sean los adecuados. Por ejemplo, los marcadores con una baja tasa de evolución podrían subestimar el tamaño efectivo, mientras que aquellos que evolucionan rápidamente sobreestimarían N_e (Avise, 2000). También hay que tomar en cuenta la homoplasia, ya que alteraría los resultados de los métodos basados en coalescencia o para el caso del tamaño del vecindario, sobreestimaría el flujo génico en la población (Cuenca *et al.*, 2003).

Por lo anterior, es necesario considerar las características de los marcadores moleculares a utilizar, ya sea su tasa de mutación, tasa de recombinación, si pueden estar sujetos a selección natural, etc., ya que estos elementos pueden afectar las estimaciones y por lo tanto deben ser considerados de manera explícita en los modelos (Nordborg y Innan, 2002).

Aplicaciones

A pesar de la vaga definición del tamaño efectivo de una población existen numerosas aplicaciones de este concepto en la literatura. Entre las más im-

portantes actualmente están la filogeografía, la epidemiología y la biología de la conservación (véase el capítulo 8 de este libro).

Filogeografía

La filogeografía es una disciplina que se enfoca en explicar los patrones de distribución geográfica de la variación genética en las poblaciones, como producto de procesos históricos y contemporáneos (Cruzan y Templeton, 2000). La estructuración genética de las poblaciones, y a la larga, la formación de nuevos linajes evolutivos, es producto de los patrones de flujo génico, la migración, la deriva génica y la dinámica demográfica de las poblaciones, determinada en parte por factores internos a las poblaciones como por factores externos como cambios climáticos y geológicos (Schaal *et al.*, 1998).

Para reconstruir esta historia se construyen genealogías de genes y utilizando como marco la teoría de la coalescencia, se pueden inferir eventos demográficos pasados y recientes que dejan huella en la estructura genética de las poblaciones. Por ejemplo, los métodos filogeográficos pueden ayudar a estimar los tamaños poblacionales de una especie en el pasado y cómo fueron sus fluctuaciones, así como detectar si éstas se dieron por posibles cuellos de botella, por expansión del rango o por colonización a larga distancia, y sus consecuencias en la evolución de la especie (Nichols *et al.*, 2002; Cruzan, 2001; Vilà *et al.*, 1999; véase el capítulo 14 de este libro).

Cabe mencionar que el empleo de marcadores haploides es de gran utilidad, ya que su N_e es la mitad de la de los genes nucleares y presentan mayor resolución para detectar cambios poblacionales (Schaal *et al.*, 1998; Cruzan y Templeton, 2000).

Epidemiología

Tradicionalmente, la demografía histórica se ha aplicado al estudio de poblaciones de mamíferos, peces, reptiles y algunas especies de plantas, pero recientemente se ha considerado como una herramienta muy útil en el control de enfermedades porque permite conocer la dinámica poblacional de los patógenos a lo largo de la historia (Hughes y Verra, 2001). La utilidad de estos métodos se basa en que si se puede determinar que un patógeno es descendiente de un solo o pocos ancestros, tendrá muy poca variación genética y por lo tanto menos variación preexistente para responder a vacunas y tratamientos, mientras que si ha tenido tamaños efectivos poblacionales muy

grandes en el pasado, tendrá mayor variación y por lo tanto mayor base de acción para la selección natural (véase el capítulo 1 de este libro). Estos dos escenarios requieren estrategias de control de la enfermedad con distintos grados de complejidad (Hughes y Verra, 2001).

Un enfoque novedoso del cálculo del tamaño efectivo se ha aplicado al estudio de la evolución de la fase asintomática del SIDA. Se ha demostrado que existe una correlación negativa entre el tamaño efectivo poblacional, dentro de un paciente, del virus VIH-1 y la tasa de evolución de las secuencias del virus, como lo predice la teoría casi neutral o de las mutaciones ligeramente deletéreas. Bajo este modelo la tasa de fijación de los alelos depende del tamaño efectivo poblacional y de los coeficientes de selección, por lo que si la mayoría de las mutaciones son ligeramente deletéreas, los coeficientes de selección serán negativos en su mayoría, siendo entonces la selección negativa la que dirige la evolución viral a tamaños poblacionales altos. Pero esta correlación podría también ser producto de la reacción del sistema inmune del paciente, ya que cuando el sistema inmune es muy fuerte, las secuencias virales estarán bajo una presión de selección muy fuerte, lo que reducirá N_e pero aumentará su tasa de evolución. El entendimiento de este fenómeno llevaría a comprender mejor cuáles son los factores que determinan el desarrollo rápido de la enfermedad en unos pacientes, y una fase asintomática muy larga en otros (Seo *et al.*, 2002).

Conclusión

El concepto de tamaño efectivo de la población, a pesar de ser un poco vago y a veces hasta oscuro, ha resultado muy útil en la biología, no sólo por sus aplicaciones sino también por su papel en el desarrollo de modelos teóricos como la coalescencia o la teoría neutral de evolución molecular, que actualmente constituyen la base de disciplinas como la filogeografía y la sistemática molecular que representan nuevas formas de contestar las preguntas clásicas de la biología evolutiva.

Es necesario el desarrollo de más y mejores métodos para calcular el tamaño efectivo poblacional, sobre todo que incorporen con mayor flexibilidad las características de las especies en su estado natural, relajando los supuestos y utilizando métodos con mayor poder estadístico que puedan lidiar con la incertidumbre de más parámetros. Esto es aplicable principalmente a los métodos genéticos, porque la inferencia del tamaño efectivo de la población es indirecta y depende mucho del modelo en el cual se basa. Por otra par-

te los estimados ecológicos requieren de mucha más información que los métodos genéticos, aunque arrojan resultados más exactos para obtener el tamaño efectivo instantáneo de la población. La desventaja de estos métodos contra los estimados genéticos es que se requiere un conocimiento mucho más completo de la biología de la especie y de datos demográficos confiables, lo que no siempre está disponible, especialmente en especies con tiempos generacionales muy largos. Por otra parte, el inconveniente de los estimados genéticos son sus incertidumbres muy altas, sobre todo si para calcularlos se tuvo que calcular muchos parámetros indirectamente. Sin embargo, en algunos casos, como la demografía histórica, son la única alternativa.

Bibliografía

Amos, W. y J. Harwood. 1998. Factors affecting levels of genetic diversity in natural populations. *Phylosophical Transactions of the Royal Society of London* 353:177-186.

Antolin, M.F., B. Van Horne, M.D. Berger Jr., A. K. Holloway, J.L. Roach y R.D. Weeks Jr. 2001. Effective population size and genetic structure of Piute ground squirrel (*Spermophilus mollis*) population. *Canadian Journal of Zoology* 79:26-34.

Avise, J. 2000. *Phylogeography. The history and formation of species*. Harvard University Press.Cambridge Mass.

Barrowclough, G.F. y R.F. Rockwell. 1993. Variance in lifetime reproductive success: estimation based on demographic data. *American Naturalist* 141:281-295.

Beerli, P. y J. Felsenstein. 1999. Maximum-likelihood estimation of migration rates and effective population numbers in two populations using a coalescent approach. *Genetics* 152:763-773.

Berthier, P., M. A. Beaumont, J.M. Cornuet y G. Luikart. 2002. Likelihood-based estimation of the effective population size using temporal changes in allele frequencies: a genealogical approach. *Genetics* 160:741-751.

Charlesworth, B. 2001. The effect of life history and mode of inheritance on neutral genetic variability. *Genetic Research* 77:153-166.

Cherry, J.L. 1998. Should we expect substitution rate to depend on population size? *Genetics* 150:911-919.

Crow, J.F. y M. Kimura. 1970. *An introduction to population genetic theory*. Harper & Row. Nueva York.

Cruzan, M.B. 2001. Population size and fragmentation thresholds for the maintenance of genetic diversity in the herbaceous endemic *Scutellaria montana* (Lamiaceae). *Evolution* 55:1569-1580.

Cruzan, M.B. y A.R. Templeton. 2000. Paleoecology and coalescence: phylogeographic analysis of hypotheses from the fossil record. *TREE* 15:491-496.

Cuenca, A., A. Escalante y D. Piñero. 2003. Long-distance colonization, isolation by distance, and historical demography in a relictual mexican pinyon pine (*Pinus nelsonii* Shaw) as revealed by paternally inherited genetic markers. *Molecular Ecology* 12:2087-2097.

Drummond, A.J., G.K. Nicholls, A.G. Rodrigo y W. Solomon. 2002. Estimating mutation parameters, population history and genealogy simultaneously from temporally spaced sequence data. *Genetics* 161:1307-1320.

Eguiarte, L.E., A. Búrquez, J. Rodríguez, M. Martínez-Ramos, J. Sarukhán y D. Piñero. 1993. Direct and indirect estimates of neighborhood and effective population size in a tropical palm, *Astrocaryum mexicanum*. *Evolution* 47:75-87.

Furlong, R.F. y J.F.Y. Brookfield. 2001. Inference of past population expansion from the timing of coalescence events in a gene genealogy. *Journal of theoretical biology* 209:75-86.

Futuyma, D.J. 1986. *Evolutionary biology*. Sinauer. Sunderland.

Gillespie, J.H. 1998. *Population genetics: a concise guide*. Johns Hopkins University Press. Baltimore.

Hedrick, P.W. 2000. *Genetics of populations*. Jones & Bartlett Publishers. Boston.

Hill, W.G. 1972. Effective size of populations with overlapping generations. *Theoretical Population Biology* 3:278-289

Hughes, A.L. y F. Verra. 2001. Very large long-term effective population size in the virulent human malaria parasite *Plasmodium falciparum*. *Proceedings of the Royal Society of London Biological Sciences* 268:1855-1860.

Kawata, M. 2001. The influence of neighborgood size and habitat shape on the accumulation of deleterious mutations. *Journal of Theoretical Biology* 211:187-199.

Kuhner, M.K., J. Yamato y J. Felsenstein. 1998. Maximum likelihood estimation of population growth rates based on the coalescent. *Genetics* 149:429-434.

Laporte, V. y B. Charlesworth. 2002. Effective population size and population subdivision in demographically structures populations. *Genetics* 162:501-519.

Merrel, D. J. 1981. *Ecological genetics*. Longman. Londres.

Nei, M. y Y. Imaizumi. 1966. Effective population size with overlapping generations. *Heredity* 21:344

Nichols, R.A., M.W. Bruford y J.J. Groombridge. 2001. Sustaining genetic variation in a small population: evidence from the Mauritius kestrel. *Molecular Ecology* 10:593-602.

Nordborg, M. y H. Innan. 2002. Molecular population genetics. *Current Opinion in Plant Biology* 5:69-73.

Nunney, L. 1993. The influence of mating system and overlapping generations on effective population size. *Evolution* 47:1329-1241.

———.2002. The effective size of annual plant populations: the interaction of a seed bank with fluctuating population size in mainaining genetic variation. *American Naturalist* 160:195-204.

Ohta T. 1992. The nearly neutral theory of molecular evolution. *Annual Review of Ecology and Systematics* 3:263-286.

Orive, M. 1993. Effective population size in organisms with complex life-histories. *Theoretical Population Biology* 44:316-340.

Schaal, B.A., D.A. Hayworth, K.M. Olsen, J.T. Rauscher y W.A. Smith. 1998. Phylogeographic studies in plants: problems and prospects. *Molecular Ecology* 7:465-474.

Seo, T., J.L. Thorne, M. Hasegawa y H. Kishino. 2002. Estimation of effective population size of HIV-1 within a host: a pseudomaximum-likelihood approach. *Genetics* 160:1283-1293.

Slatkin, M. y N.H. Barton. 1989. A comparison of three indirect methods for estimating average levels of gene flow. *Evolution* 43:1349-1368.

Slatkin, M. y R.R. Hudson. 1991. Pairwise comparisons of mitochondrial-DNA sequences in stable and exponentially growing populations. *Genetics* 129:555-562

Storz, J.F., U. Ramakrishnan y S.C. Alberts. 2001. Determinants of effective population size Implicaciones evolutivas for loci with different modes of inheritance. *Journal of Heredity* 92:497-502.

Thévenon, S. y D. Couvet. 2002. The impact of inbreeding depression on population survival depending on demographic parameters. *Animal Conservation* 5:53-60.

Tremblay, R.L. y J.D. Ackerman. 2001. Gene flow and affective population size in *Lepanthes* (Orchidaceae): a case for genetic drift. *Biological Journal of the Linnean Society* 72:47-62.

Vilà, C., I.R. Amorim, J.A. Leonard, D. Posada, J. Castroviejo, F. Petrucci-Fonseca, K.A. Crandall, H. Ellegren y R.K. Wayne.1999. Mitochondrial DNA phylogeography and population history of the grey wolf *Canis lupus. Molecular Ecology* 8:2089-2103.

Williamson, E.G. y M. Slatkin. 1999. Using maximum likelihood to estimate population size from temporal changes in allele frequencies. *Genetics* 152:755-761.

Wright, S.1930. Evolution in mendelian populations. *Genetics* 16:97-158.

———. 1943. Isolation by distance. *Genetics* 28:114-138.

———. 1969. *Evolution and the genetics of populations.* Vol. 2: The theory of gene frequencies. University of Chicago Press. Chicago.

Capítulo 4

Reconstrucción de la historia de cambio de los caracteres

León Martínez Castilla

Una de las concepciones fundamentales de la biología es que los seres vivos contemporáneos están conectados entre sí por relaciones de parentesco y que descienden de ancestros comunes más o menos alejados en el tiempo. Intuitivamente, entre más reciente haya sido el ancestro común de dos seres vivos (u otras entidades biológicas como poblaciones, especies, taxa superiores) mayor debe ser la similitud entre estos organismos. La metáfora clásica del proceso evolutivo de separación y diferenciación de las entidades biológicas es un árbol cuyas ramas (es decir, los linajes de organismos) se bifurcan siempre. Recientemente esta metáfora ha debido modificarse un poco para incorporar el hecho de que a veces los linajes de seres vivos, o por lo menos de aquellos de sus componentes que usamos para inferir sus parentescos, se fusionan entre sí en lugar de bifurcarse siempre (véanse los capítulos 9 y 13 de este libro).

El propósito de la reconstrucción filogenética es tratar de inferir las relaciones de parentesco o de ancestría-descendencia de un conjunto de datos representativos de entidades biológicas. A menudo estos datos son secuencias de nucleótidos o aminoácidos, como las que se considerarán en este capítulo, y además de permitir reconstruir la filogenia de las especies pueden usarse para establecer cómo están relacionadas diferentes poblaciones de una misma especie o incluso diferentes miembros de una familia multigénica. Los niveles de organización a los que pertenecen las entidades que se pueden estudiar

en un marco conceptual filogenético incluyen no sólo las especies, sino las poblaciones, los genes, los organismos individuales, etc., que en general se denominan, en la jerga filogenética, unidades taxonómicas operacionales (OTUs por sus siglas en inglés).

Una vez que se ha reconstruido la filogenia de las entidades estudiadas, ésta se puede usar para hacer pruebas de hipótesis en una gran variedad de áreas de la ecología, la evolución y la biomedicina. Por ejemplo, la reconstrucción filogenética se ha convertido en un componente indispensable de los estudios basados en el método comparativo que tratan de establecer si una determinada característica de un organismo es producto de la selección natural o de otros proceso evolutivos. Las reconstrucciones filogenéticas también son necesarias para conocer e interpretar la secuencia de cambios que ha sufrido una característica a través de la historia del linaje de los organismos que la poseen; determinar si un sistema de organismos relacionados ecológicamente constituye un ejemplo pertinente de coevolución; hacer estudios de biogeografía y filogeografía (véanse los capítulos 13, 14 y 15 de este libro); determinar las relaciones de ortología y paralogía entre genes cuando se comparan genomas de diferentes especies (Mouse Genome Sequencing Consortium, 2002); estimar tiempos de divergencia entre grupos de organismos (Aris-Brosou y Yang, 2002); reconstruir proteínas ancestrales (Chang *et al*., 2002); encontrar residuos o aminoácidos de una proteína que se encuentran bajo el efecto de la selección natural y por lo tanto parecen ser críticos para la adecuación del organismo del que son parte (Bush *et al.*, 1999; Nielsen y Huelsenbeck, 2002; Anisimova *et al*., 2002); detectar puntos de recombinación (p. ej., se han utilizado métodos filogenéticos bayesianos para determinar qué cepas del virus de inmunodeficiencia humana surgieron por recombinación [Suchard *et al*., 2002; véase el capítulo 9 de este libro]); identificar mutaciones que probablemente están asociadas con enfermedades (Fleming *et al*, 2001); y determinar la identidad de patógenos emergentes (Rota *et al*., 2003; Hughes *et al.,* 1993; Anderson *et al*., 1999; Lanciotti *et al*., 1999).

Cualquier comparación de más de dos secuencias relacionadas por parentesco presupone, así sea implícitamente, una filogenia subyacente. Para algunos análisis las comparaciones entre pares de secuencias son suficientes (por ejemplo, BLAST), pero incluso estos algoritmos pueden volverse mucho más poderosos si se considera la información de más de dos secuencias simultáneamente (por ejemplo, PSI-BLAST). Tomando en cuenta el potencial de la genómica comparada y su relativo bajo costo, incluso investigadores que no tienen interés en la evolución de los organismos o de los genes *per se* podrían

estar al tanto de las aproximaciones filogenéticas. Sin embargo, la multitud de métodos contrastantes y la diversa terminología de la sistemática puede resultar intimidante para quien busca simplemente aplicar los métodos para contestar preguntas de tipo práctico, de manera que vale la pena familiarizarse con sus detalles.

Árboles, genes y organismos

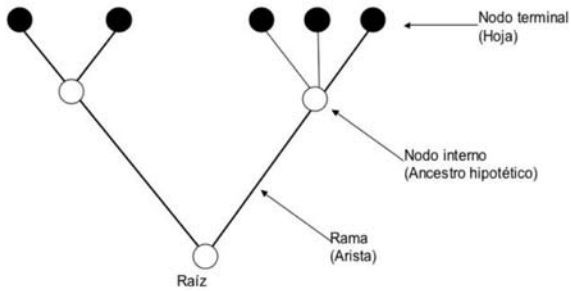
La noción central del quehacer de la sistemática es el **árbol filogenético**. Un árbol es un grafo, una estructura matemática que describe cómo se relacionan entre sí, por cualesquiera tipos de relación, los diferentes elementos de un conjunto. Un árbol filogenético es un grafo que describe las relaciones de parentesco, ancestría y descendencia, entre un conjunto de genes, de organismos, de especies o de taxa superiores. Las relaciones de parentesco se infieren a partir de un conjunto de características representativas que pueden ser más o menos compartidas entre las entidades biológicas cuyas relaciones se intenta dilucidar, lo que, a grandes rasgos, permitirá concluir que están emparentadas en mayor o menor grado. En este capítulo se parte del supuesto que las características representativas son secuencias, ya sea de ácidos nucleicos o de los aminoácidos que constituyen proteínas.

Podemos imaginar que las secuencias son como hojas que se encuentran en las puntas del árbol filogenético. Como se supone que las entidades biológicas están emparentadas a través de ancestros, las ramas conectan a las secuencias, pero indirectamente, siempre a través de una secuencia ancestral hipotética, generalmente no observable (excepto en casos de evolución extremadamente rápida, como la de los virus que infectan a un hospedero, en donde es válido postular que algunos virus muestreados son descendientes de otros observados un poco antes durante un estudio). Aquí usaremos los términos árbol, árbol filogenético y filogenia de forma intercambiable.

El lenguaje de los árboles

Formalmente, un árbol consiste en **nodos** conectados por **ramas** (también llamadas bordes o aristas). Los **nodos terminales** (también llamados OTUs, hojas o taxa terminales) representan secuencias u organismos de los que tenemos datos, puede tratarse de entidades biológicas extintas o vivientes. Los **nodos internos** representan a los ancestros hipotéticos y el ancestro de todas las secuencias comprendidas en el árbol es la **raíz** del árbol (figura 1).

Figura1. Un árbol simple y sus términos asociados

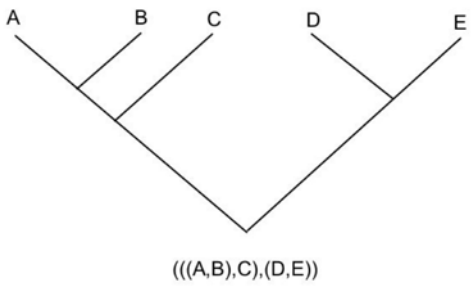


Los nodos y ramas de un árbol contienen diferentes clases de información. Por ejemplo, algunos métodos de reconstrucción de filogenias buscan reconstruir las características de cada ancestro hipotético, otros tratan de estimar el grado de divergencia que ha ocurrido entre cada nodo del árbol, que puede ser representado como **longitudes de ramas**. El número de ramas que llegan o salen de un nodo se denomina el **grado** de ese nodo. Si un nodo tiene un grado igual o superior a tres entonces es una politomía, mientras que un árbol que no tiene politomías se considera completamente resuelto. Un árbol que tiene politomías puede representar diferentes situaciones. Por un lado, es posible que los diferentes linajes que se derivaron del nodo politómico se separaran al mismo tiempo en vez de secuencialmente o al menos tan rápido que nuestros métodos no nos permiten distinguir el orden específico de separación. Alternativamente, la politomía puede indicar que la información que tenemos no permite decidir por uno de los órdenes de ramificación posibles, quizá porque diferentes partes de nuestro conjunto de datos se contradicen entre sí. Al primer caso se le llama **politomía dura**, mientras que al segundo se le llama **politomía blanda**.

Los árboles filogenéticos pueden representarse mediante una notación compacta, basada en paréntesis anidados, además de la representación gráfica convencional. En la figura 2 se puede ver un ejemplo de esta notación compacta, llamada formato Newick o New Hampshire.

Los árboles pueden contener diferentes tipos de información acerca de las relaciones que hay entre sus OTUs. El tipo de árbol que contiene la información más simple es el **cladograma**, que simplemente muestra qué tan reciente es la ancestría común entre los nodos terminales. La figura 2 indica que las

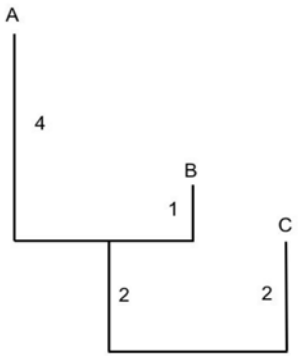
Figura 2. Un árbol y su representación en notación compacta



secuencias A y B comparten un ancestro más cercano de lo que cualquiera de las dos comparte con C. Nótese que se podría hacer girar las ramas alrededor del nodo interno del que surgen y la información sobre qué tan reciente es la ancestría común de las OTUs, lo que también se conoce como la topología de un árbol, no cambiaría.

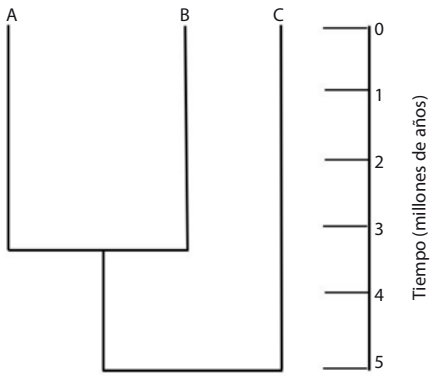
El **árbol aditivo** es otro tipo de árbol que contiene información adicional acerca de la cantidad de cambio evolutivo que separa a las OTUs, asignando diferentes longitudes a las ramas del árbol. Por ejemplo, en la figura 3 se supone que la secuencia A adquirió cuatro substituciones desde que su linaje se separó del de la secuencia B. A los árboles que contienen este tipo de información también se les llama árboles métricos o filogramas.

Figura 3. Un árbol aditivo que representa la cantidad de cambio evolutivo que ha tenido lugar desde la divergencia de las secuencias A, B y C. Los números representan las sustituciones de una secuencia a otra



Los **árboles ultramétricos** son un tipo especial de árbol aditivo en el que los nodos terminales están todos a la misma distancia de la raíz del árbol y sirven para representar el tiempo que separa a los grupos de secuencias.

Figura 4. Un árbol ultramétrico que muestra el tiempo de divergencia en millones de años entre las secuencias A, B y C



Árboles enraizados y no enraizados

Los árboles pueden estar **enraizados** (o polarizados) o no. Un árbol enraizado contiene un nodo del cual descienden, en última instancia, todos los demás nodos, de manera que tiene una direccionalidad que corresponde al tiempo evolutivo: entre más cerca esté un nodo del nodo raíz, más viejo será. De esta manera, los árboles enraizados nos permiten definir relaciones de ancestría-descendencia: si dos nodos están conectados por una rama o serie de ramas, aquel que esté más cerca de la raíz será ancestro del que esté más cerca de la punta. En un árbol que no está enraizado no podemos definir el orden de aparición de los eventos evolutivos, por lo que su utilidad es más limitada en el estudio de la evolución. Algunos de los métodos de reconstrucción filogenética generan árboles que no están enraizados y por lo tanto no hacen ninguna diferencia entre los árboles de la figura 5. El número de posibles árboles no enraizados NoE_n para n secuencias es:

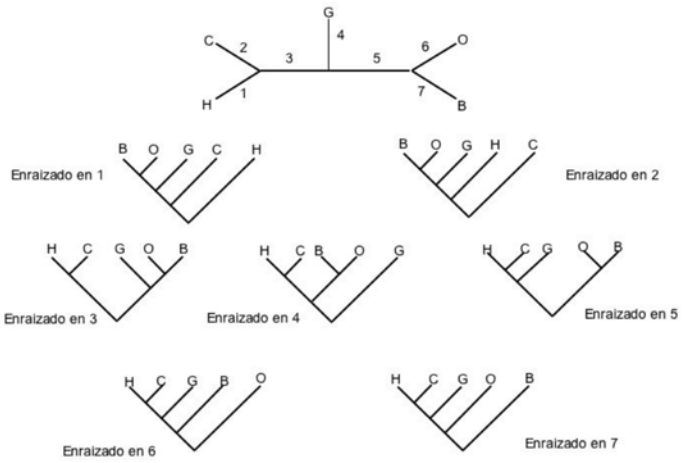
$$NoE_n = (2n\text{-}5)\ (2n\text{-}7)\ldots\ (3)\ (1)$$

para $n \geq 2$. El número de posibles árboles enraizados E_n para $n \geq 3$ es

$$E_n = (2n\text{-}3)\,(2n\text{-}5)\ldots\,(3)\,(1) = (2n\text{-}5)\,NoE_n$$

como se puede ver, el número de árboles no enraizados posibles, y aún más el de los enraizados crece muy rápidamente con el aumento del número de secuencias. Por ejemplo, para 20 secuencias hay 8 200 794 532 637 891 559 000 árboles no enraizados posibles. Esto plantea una dificultad para los métodos de reconstrucción de filogenias que tratan de estimar cuál de todos los árboles posibles representa la estimación óptima de la filogenia real.

Figura 5. Un árbol enraizado para el ser humano (H), el chimpancé (C), el orangután (O), el gorila (G) y el gibón (G). Los siete árboles enraizados cada uno en una rama diferente corresponden al árbol no enraizado



Reconstrucción de la historia del cambio de los caracteres

Si bien es muy interesante para nosotros conocer cuáles son las relaciones filogenéticas entre las entidades biológicas, a menudo esto es sólo un primer paso en una investigación que tiene otros objetivos. Por ejemplo, podemos

estar interesados en saber si determinado cambio evolutivo, como sería una sustitución nucleotídica, ocurrió en alguno de los linajes que estamos estudiando; quizá también podría interesarnos saber si el hecho de que un mismo aminoácido se encuentre en la misma posición en dos proteínas es consecuencia de procesos de evolución independientes o bien se explica porque el ancestro común de esas macromoléculas también poseía ese aminoácido en esa posición. Para poder abordar cuestiones de este tipo es necesario reconstruir la historia del cambio de los caracteres.

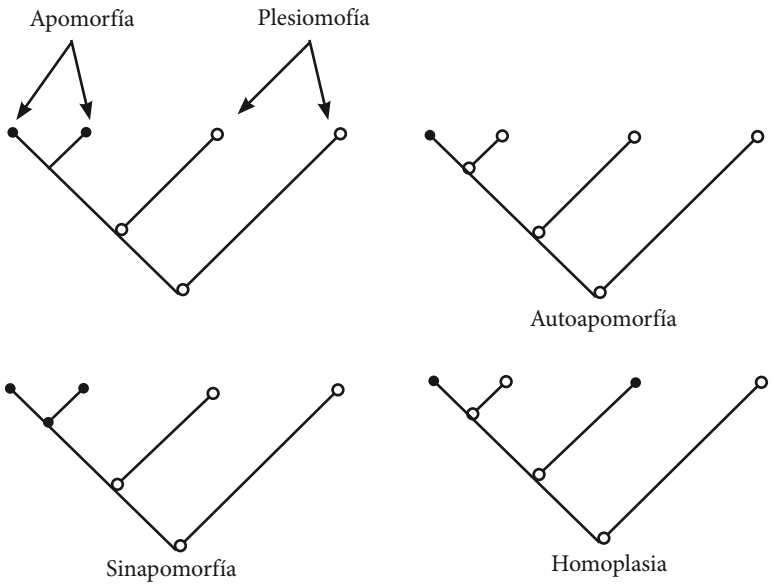
Si planteamos que determinados caracteres (secuencias de nucleótidos, de aminoácidos o incluso observaciones sobre la morfología de diferentes organismos) están conectados por un árbol filogenético, podemos plantear que algunos de ellos son ancestrales y otros son derivados. Es decir, postulamos que en los organismos actuales hay caracteres que se parecen a los de sus ancestros lejanos (caracteres ancestrales o *primitivos*) y que hay otros caracteres que en los organismos actuales se encuentran como una versión modificada de los que tenían sus ancestros (caracteres derivados).

Si una macromolécula tiene en una posición dada el mismo nucleótido o el mismo aminoácido que el ancestro común de todas las otras macromoléculas con las que la estamos comparando, entonces diríamos que ese caracter se encuentra en un estado ancestral o plesiomórfico. En cambio, si en esa posición hay un nucleótido o un aminoácido diferente del que tenía el ancestro, el estado de ese caracter para esa secuencia sería derivado o apomórfico. Los caracteres apomórficos que sólo están presentes en una de las secuencias estudiadas se denominan autoapomorfías (formas derivadas aisladas) mientras que los estados de caracter derivados que estén presentes en varias secuencias se llaman sinapomorfías (formas derivadas compartidas).

Cuando dos o más secuencias tienen un estado de caracter idéntico (por ejemplo el mismo nucleótido en la misma posición), la similitud puede explicarse porque las secuencias heredaron ese estado para ese caracter de su ancestro común, en cuyo caso se trataría de una homología. También es posible que la similitud se deba a que cada secuencia adquirió independientemente ese estado de caracter, sin que el caracter tuviera ese estado en el ancestro común, en cuyo caso se trataría de una homoplasia. Sólo la homología refleja adecuadamente la ancestría común de los organismos. La homoplasia no es un buen indicador de las relaciones filogenéticas porque la similitud homoplásica no refleja la ancestría común (Page y Holmes, 1998; Morrone, 2000).

La homoplasia puede deberse a que las secuencias comparadas hayan adquirido el estado de caracter independientemente a partir de la misma

Figura 6. Árboles que muestran la terminología usada para describir los diferentes patrones de estados de caracter ancestrales (○) y derivados (●); tomado de Page y Holmes (1998)



condición ancestral (en cuyo caso se habla de paralelismo) o a partir de diferentes condiciones ancestrales (y entonces se habla de convergencia). La homoplasia también puede deberse a la pérdida secundaria de una característica derivada (es decir, perder esa característica después de haberla ganado), lo que hace que el estado de caracter *parezca* ancestral (Page y Holmes, 1998).

Ancestros

Toda filogenia de entidades biológicas presupone la existencia de ancestros –entidades de cuyos descendientes obtenemos los caracteres que usamos para reconstruir las relaciones filogenéticas. Estos ancestros están representados en los nodos internos de los árboles y son generalmente hipotéticos, aunque la obtención de ADN fósil y los estudios de evolución experimental o de seguimiento a largo plazo de infecciones por bacterias y virus plantean la

problemática de construir filogenias en las que se puedan incluir en los nodos internos caracteres observados actuales.

Medidas de distancia

Existen varias medidas de la divergencia que hay entre dos grupos de datos que pueden usarse para estimar el número de cambios evolutivos que ha ocurrido entre dos secuencias biológicas desde que existió su último ancestro común (distancia evolutiva). Esta distancia puede representarse en los árboles filogenéticos, lo que ha motivado la creación de una serie de métodos que buscan convertir estas distancias pareadas en árboles que relacionen más de dos secuencias.

Para que una medida de distancia pueda usarse para reconstruir una filogenia debe tener varias características. La primera es que debe de tratarse de un *métrico* ¿Qué es un métrico? Supongamos que *d(a,b)* es la distancia entre las secuencias *a* y *b*, entonces, *d* será un métrico si cumple con estas condiciones:

1) $d(a,b) \geq 0$. Esto se conoce como no-negatividad.
2) $d(a,b) = d(b,a)$. Es decir, simetría.
3) $d(a,c) \leq d(a,b) + d(b,c)$. Esto se llama *desigualdad triangular* porque equivale a asegurar que las distancias entre las tres secuencias *a*, *b* y *c* se pueden representar geométricamente como un triángulo.
4) d(a,b) = 0 si y sólo si a=b. Esto se conoce como "distintividad" e indica que secuencias que no sean idénticas deben de estar a una distancia diferente de cero.

   Las condiciones 1, 2 y 4 generalmente se cumplen para datos obtenidos directamente de secuencias moleculares, pero medidas indirectas como la hibridación ADN-ADN o mediciones de reacción inmunológica cruzada no siempre obedecen estas condiciones, particularmente la número 2 (Page y Holmes , 1998).

   Si, además, una distancia satisface el criterio de que
5) $d(a,b) \leq \max[d(a,c), d(b,c)]$ entonces la distancia, además de ser un métrico, es *ultramétrica*.

   Esto implica que las dos distancias más grandes son iguales y definen un triángulo isósceles. Desde el punto de vista evolutivo, las distancias ultramétricas implican que existe una constante de divergencia entre secuencias; que las distancias que sean más similares entre sí también serán las que

estén más cercanamente emparentadas y que la distancia total entre la raíz del árbol y cualquiera de sus puntas es constante.

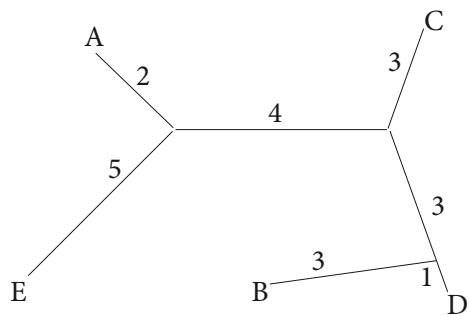
Para que una distancia sea una medida válida de cambio evolutivo debe también la condición de aditividad:

6) $d(a,b) + d(c,d) \leq \max[d(a,c)+d(b,d), d(a,d)+d(b,c)]$
lo que es equivalente a exigir que de las tres sumas
$d(a,b)+d(c,d)$, $d(a,c)+d(b,d)$ y $d(a,d)+d(b,c)$
las dos más grandes sean iguales.

Para comprender más claramente la aditividad, considérese a *D*, una matriz simétrica n×n donde todos los valores en la diagonal son iguales a cero y todos los demás valores son estrictamente positivos. Sea *A* un árbol a cuyas ramas están asociados valores, que tiene por lo menos *n* nodos y en donde los nodos terminales de *A* están etiquetados con las hileras de la matriz *D*. El árbol *A* se considera un árbol aditivo para la matriz *D* si para cada par de nodos etiquetados *i,j*, la ruta que los conecta tiene un peso (o una distancia) exactamente igual a D_{ij}.

Por lo tanto D_{ij} representa la suma de las longitudes de ramas en el camino que hay entre *i* y *j* en el árbol *A*.

	A	B	C	D	E
A	0	12	9	10	7
B		0	6	4	15
C			0	7	1
D				0	13
E					0



Árboles de consenso

Si quisiéramos integrar dos o más reconstrucciones filogenéticas que se refieren al mismo grupo de taxa o de OTUs, pero que se generaron con métodos diferentes o con juegos de datos diferentes (por ejemplo, diferentes secuencias genéticas; con secuencias moleculares por un lado y con datos morfológicos por el otro; con diferentes juegos de *bootstraps* (véase la sección Evaluación

estadística de árboles) o cuando se trata de reunir la información de un grupo de árboles igualmente parsimoniosos) podríamos usar los árboles de consenso, que muestran la información que hay en común en un conjunto de árboles. Cuando existe conflicto entre dos o más árboles filogenéticos, el árbol de consenso se representa como una politomía (una politomía blanda, pues existe incertidumbre sobre cuál es el orden correcto de ramificación).

Existen diferentes criterios sobre cuál es el aspecto relevante que se trata de plasmar en el árbol de consenso. Por ejemplo, en un consenso estricto sólo se incluyen aquellos agrupamientos filogenéticos que están presentes en todos los árboles que se están tratando de integrar.

Es posible que algunos agrupamientos estén presentes en la mayoría de los árboles considerados pero no en todos, información que podría perderse en el árbol de consenso estricto. Para rescatarla podríamos hacer un consenso de mayoría en el cual estaría representado cualquier agrupamiento que se encontrara en más de la mitad de los árboles considerados (Page y Holmes, 1998).

Recontrucción de árboles filogenéticos

La reconstrucción de árboles filogenéticos con datos moleculares intenta extraer la información evolutiva que se encuentra en los genomas y proteomas de los organismos, pero ¿en qué parte de dichos genomas o proteomas se encuentra la información más adecuada para construir una filogenia? Obviamente, la respuesta dependerá del motivo por el que queremos reconstruir la filogenia. Por ejemplo, no recurriríamos al mismo tipo de secuencias si tratáramos de dilucidar las relaciones entre grupos de organismos que se formaron en las etapas tempranas de la vida, que si nos interesáramos por las relaciones entre las variedades que pertenecen a especies de origen relativamente reciente como el maíz, o si nos interesara hacer un diagnóstico de las variedades patógenas de una bacteria.

Podría ser también que no quisiéramos reconstruir una filogenia a partir de datos moleculares, sino que nos interesa la historia de las moléculas en sí, y no buscamos usarlas como evidencia sobre las relaciones genealógicas de los organismos estudiados. Es decir, puede interesarnos más la evolución de las moléculas que la historia de los organismos de donde provienen (Morrone, 2002). Para decirlo muy concretamente, no todos los genes sirven para la reconstrucción filogenética, ya que la información evolutiva no está distribuida uniformemente en las moléculas informacionales. Para entender qué criterios hay que tomar en cuenta al escoger la secuencia o grupos de secuencias para un análisis de filogenia molecular, haremos un

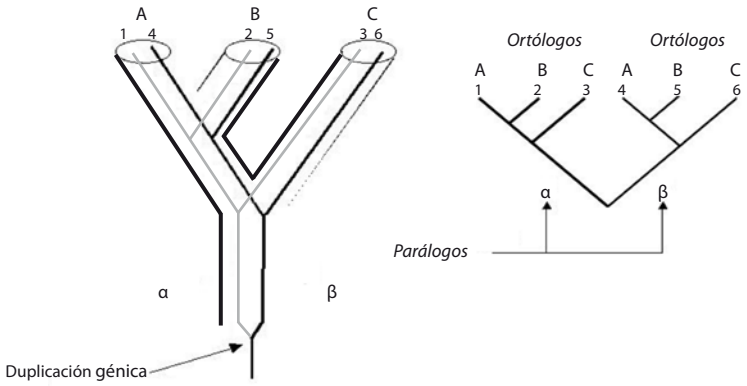
breve repaso sobre cómo la información filogenética puede conservarse o perderse en las moléculas.

Árboles de genes y árboles de especies

A primera vista podría creerse que existe una relación biunívoca entre las filogenias de genes y las filogenias de organismos, de manera que las primeras nos dan las segundas. Sin embargo, hay varias razones por las que esto no es necesariamente así. Una es que la duplicación de genes puede hacer que una misma especie tenga varias secuencias distintas pero relacionadas entre sí.

En el ejemplo de la figura 7, cada una de las tres especies A, B y C tiene dos copias, α y β, de un gen. Una filogenia para los seis genes nos permite recuperar la filogenia organismal correcta ((A,B),C) a partir tanto de los genes α como de los β. Sin embargo, si por desgracia secuenciáramos únicamente los genes 1, 3 y 5 y no supiéramos que son parte de un árbol de genes más grande, infeririámos que el árbol organismal es ((A,C),B) porque el gen 3 de la especie C está relacionado más cercanamente con el gen 1 de la especie A de lo que está relacionado con el gen 5 de la especie B, aunque la especie B está más relacionada con la especie A que con la especie C.

Figura 7. Filogenia para tres especies A, B y C y seis genes que se derivan de una duplicación que originó a dos clados parálogos de genes α y β. Los genes α1-3 son ortólogos entre sí, al igual que los son los genes β 4-6; sin embargo, cada gen α es parálogo con cada gen β puesto que lo que los separó fue un evento de duplicación y no un evento de especiación (Page y Holmes, 1998)



Esto ilustra que la filogenia organismal sólo se puede inferir correctamente si tenemos un juego completo de genes o si nos limitamos a usar genes que no hayan sufrido duplicaciones. Es decir, necesitamos genes que sean ortólogos entre sí. Dos genes homólogos son *ortólogos* si su ancestro común más reciente no sufrió una duplicación génica, en caso contrario se dice que son *parálogos* entre sí. En la figura 7, los genes 1-3 son ortólogos entre sí, al igual que lo son los genes 4-6, pero cualquier par de genes α y β son parálogos entre sí.

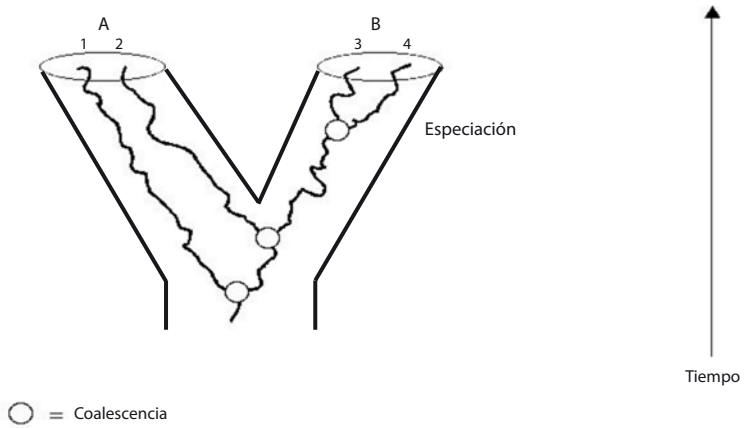
Otra complicación surge cuando se pretende conocer el grado de asociación entre dos filogenias –por ejemplo una de organismos y otra de genes; una de especies de parásitos y otra de sus hospederos; o una filogenia de organismos confrontada a un esquema de la historia de los eventos geológicos que han afectado las regiones en que se encuentran (Morrone, 2002; Page y Holmes, 1998)– y algunos de los linajes relevantes no fueron detectados e incluidos en el estudio. Esta falta de detección puede tener razones ontológicas – los linajes faltantes no han sido detectados porque realmente desaparecieron, o un organismo ancestral perdió una de las copias de una duplicación génica y por lo tanto sus descendientes muestran, de forma derivada, una sola copia del gen –o bien, metodológicas– su ausencia se debe a que nuestro muestreo fue incompleto.

Incluso si estamos seguros de que nuestro estudio ha usado únicamente genes ortólogos entre sí, la posibilidad de que los diferentes alelos hayan sobrevivido diferencialmente puede tener como consecuencia que la filogenia de los alelos no corresponda exactamente a la filogenia de los organismos. Consideremos un par de alelos ortólogos e imaginemos que retrocedemos en el tiempo; en algún momento llegaremos a un nodo que representa a su ancestro común más reciente. Se dice que en este nodo los alelos coalescen y el tiempo que hay que retroceder desde las puntas de las ramas del árbol hasta que los alelos coalescen se llama tiempo de coalescencia.

En el ejemplo que se muestra en la figura 8 los alelos 3 y 4 tienen un punto de coalescencia que está dentro del tiempo que lleva existiendo el linaje organísmico B. Por el contrario, los alelos 1 y 2 tienen un tiempo de coalescencia más antiguo que antecede al tiempo de aparición de su linaje, es decir, son más viejos que el linaje A. Además, aunque tanto el alelo 1 como el 2 se encuentran en la misma especie, no son el pariente más cercano el uno del otro.

La presencia de un par parafilético de alelos en el linaje A puede tener consecuencias en el tiempo evolutivo. Imaginemos que poco después de la divergencia de las especies A y B, y mientras todavía existieran los alelos 1 y 2, la especie A divergiera en las especies A_1 y A_2. En este caso, la especie A_1 hereda el alelo 1 y la especie A_2 hereda el alelo 2. Dicho de otra forma, los dos

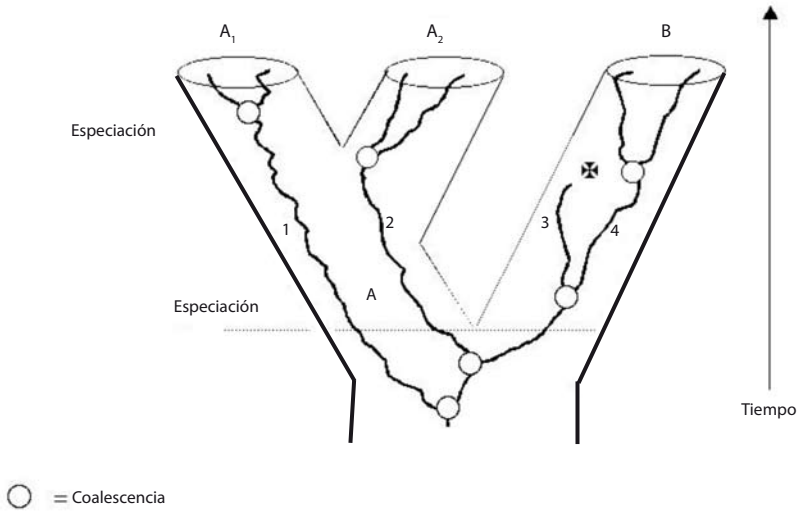
Figura 8. Un árbol de genes para cuatro alelos de un mismo gen (1-4) en dos linajes de organismos separados por especiación *A* y *B*. Los alelos 3 y 4 coalescen dentro del linaje B (es decir, se separaron después de que el linaje B se hubiera formado y son exclusivos de este linaje) pero los alelos 1 y 2 son más viejos que el linaje A. En particular, los alelos 1 y 2 no forman un grupo monofilético, el alelo 2 es más cercano a los alelos 3 y 4 que al otro alelo presente en el linaje A (el alelo 1). Los alelos 3 y 4 tienen un tiempo de coalescencia más reciente que el tiempo de aparición del linaje B (Page y Holmes,1998)



linajes de alelos, 1 y 2, fueron repartidos (*sorted*) entre los descendientes de A. Nótese que aunque las tres especies tienen juegos monofiléticos de alelos, los alelos que se encuentran en A_2 están en realidad más cercanamente emparentados con los de la especie B que con su especie hermana A_1. Si utilizáramos la filogenia de estos alelos para inferir la filogenia de las tres especies A_1, A_2 y B concluiríamos incorrectamente que el árbol de especies sería (A_1, (A_2, B)). Este ejemplo ilustra el problema de la repartición de linajes (*lineage sorting*). Si los alelos de un linaje no son monofiléticos antes de que ese linaje sufra especiación, la distribución y las relaciones de estos alelos no necesariamente reflejarán la filogenia de los organismos.

La repartición de linajes se convierte en un problema para las filogenias organismales si el tiempo necesario para que coalescan dos alelos dentro de un linaje es mayor que el intervalo entre eventos sucesivos de especiación. La figura 10 muestra una extrapolación alternativa de la figura 9, en la que la especie A se divide en dos especies hijas más tardíamente. Cuando A especia,

Figura 9. Una posible extrapolación sobre el futuro de los árboles de genes y de especies que se muestran en la figura 2. En este caso, la especie A divirgió en las especies A_1 y A_2. Los dos alelos, 1 y 2, presentes en A cuando especió, fueron heredados por A_1 y A_2 respectivamente. El alelo 3 se extinguió



sus descendientes reciben un conjunto monofilético de alelos. En este caso, la filogenia de alelos refleja fielmente la filogenia de las especies.

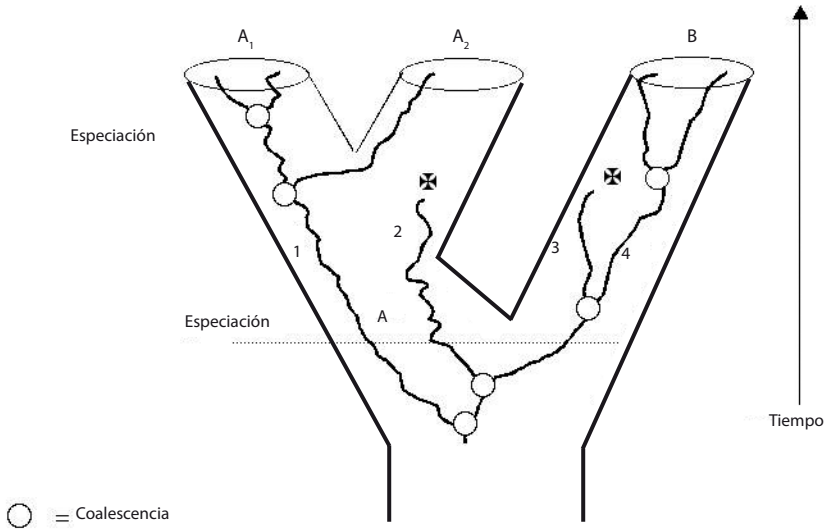
La diferencia entre la figura 9 y la 10 es el tiempo que pasa entre eventos sucesivos de especiación en el mismo linaje. Debido a una combinación de eventos fortuitos y de selección natural, los linajes de alelos podrán persistir, radiar o extinguirse. Entre más largo sea el intervalo entre los eventos de especiación, mayor es la probabilidad de que estos procesos desemboquen en linajes con un conjunto monofilético de alelos (Page y Holmes, 1998).

Redes

Hasta aquí hemos dado por sentado que la mejor manera de representar las relaciones evolutivas entre secuencias y organismos es mediante un árbol, pero ¿qué pasaría si en realidad la historia evolutiva no es un proceso de ramificación arborescente, de bifurcación y divergencia?

Por ejemplo, la misma metáfora del "árbol genealógico" es engañosa. Un árbol tiene una sola raíz y cuando sus ramas se separan no se vuelven a unir,

Figura 10. La misma situación que en la figura anterior, pero aquí el linaje A se divide más tardíamente en dos especies, para cuando el alelo 2 ya se ha extinguido. Por lo tanto, las especies A_1 y A_2 heredan un conjunto monofilético de alelos



mientras que en una genealogía, cada vez que nace un descendiente producto de un apareamiento las ramas de los padres se fusionan. En general la historia de cada gen individual puede representarse razonablemente bien como un árbol, sin embargo, cuando un gen ha sufrido recombinación, una red puede ser una mejor representación. Una red es un grafo que contiene uno o más ciclos, es decir un conjunto de nodos entre los cuales es posible trazar un camino que comience y termine en el mismo nodo sin visitar ningún otro nodo más de una vez. Dicho de otra forma, una red es un grafo en el que, para al menos un par de nodos, existe más de un camino que los conecta.

En general cuando los datos con los que se quiere reconstruir una historia evolutiva se pueden subdividir en regiones que han tenido diferentes historias, las redes pueden ser una alternativa para la representación de las relaciones filogenéticas (Page y Homes, 1998). Un caso es el de la recombinación, pero lo es también la transferencia horizontal, que provoca que el genoma de algunos organismos tenga secuencias que no contienen información histórica propia de su linaje sino, al menos en parte, de otros linajes que fueron los que "donaron" la secuencia al linaje donde se observa hoy.

Homología y alineación

Dado que los estudios de sistemática y ecología molecular son en última instancia estudios de biología comparada, es importante enfatizar que los datos que van a representar cada unidad taxonómica en un análisis filogenético deben ser homólogos a los datos de las demás unidades taxonómicas. No hacerlo sería equivalente a caer en el error proverbial de comparar peras con manzanas. Nosotros usaremos la acepción evolutiva de la palabra homología, es decir, consideraremos que dos o más estructuras o procesos son homólogos entre sí, si y solo si adquirieron su estado actual por herencia directa de su ancestro común (Page y Holmes, 1998; Morrone, 2002; Schuler, 2001; Barton, 2001). Un corolario de esta afirmación es que no hay grados de homología: los caracteres, que pueden ser génicos (moleculares), estructurales, morfológicos o funcionales (metabólicos, regulatorios, conductuales), son homólogos o no lo son. De esta manera, es válido decir que dos caracteres son 60% idénticos, pero no está justificado decir que son 60% homólogos (Fitch, 2000). Pero nótese que esta definición de homología no exige que la totalidad de los caracteres de un gen o de una proteína tenga que ser homóloga con la totalidad del otro gen o proteína. Como han señalado Fitch (2000) y Hillis (1994) un gen puede construirse a partir de dominios. Cuando esto ocurre, el término de homología no se aplica al gen completo, sino que se tiene una serie de dominios, cada uno de los cuales es parálogo a un dominio similar de un gen diferente. Hemos de reconocer que no todas las partes de un gen tienen la misma historia y que en esos casos, el gen no es la unidad a la que se aplican los términos ortología, paralogía, etc.

Usemos un caso concreto para ejemplificar esto: la enterocinasa tiene por lo menos 6 dominios, uno de los cuales se relaciona con un receptor de lipoproteína de baja densidad. Si el dominio que es homólogo al receptor de lipoproteína constituye el 20% de la enterocinasa, entonces la enterocinasa es solamente 20% homóloga al receptor de lipoproteína, independientemente de cuál sea su porcentaje de identidad. Al mismo tiempo, si este dominio en común constituyera la mitad del receptor de lipoproteína, el receptor sería 50% homólogo a la enterocinasa. Si las proteínas son de diferentes longitudes, las homologías no son las mismas en las dos direcciones. De acuerdo con Fitch (2000), ésta es la única situación en la que se justifica hablar de porcentaje de homología, aunque Hillis (1994) sugiere que hablar de homología parcial se presta a menos confusiones.

Nótese también que determinar si una característica de una entidad biológica (trátese de secuencias moleculares, de mediciones morfológicas o de patrones de restricción) es homóloga de otra requiere conocer previamente las relaciones evolutivas de las especies (u otras entidades biológicas) que tienen esas características. El lector astuto habrá notado que aquí está agazapado un problema de circularidad: queremos asegurarnos de que algunas secuencias son homólogas para poder reconstruir su filogenia, pero para saber si son homólogas necesitamos conocer sus relaciones filogenéticas. Una manera práctica de escapar de este dilema consiste en invocar un criterio independiente del de homología para establecer en un primer paso qué regiones de las secuencias se pueden considerar instancias derivadas de la misma región de un gen y por lo tanto se pueden comparar. Probablemente el criterio independiente de asignación de homología más intuitivo sea la similitud. Considérense las siguientes dos secuencias de nucleótidos:

```
Secuencia 1  ATGCGTCGTT
Secuencia 2  ATGCGTCGT
```

Claramente son muy similares, pero no son idénticas. Para compararlas podemos recurrir a una herramienta gráfica llamada matriz de puntos. Estas matrices se construyen poniendo cada una de las secuencias que se desea comparar en los ejes de un gráfico bidimensional y colocando un punto en donde las dos secuencias concuerden. Si las dos secuencias fueran idénticas, se podría dibujar una diagonal continua desde el cuadrante inferior izquierdo hasta el superior derecho.

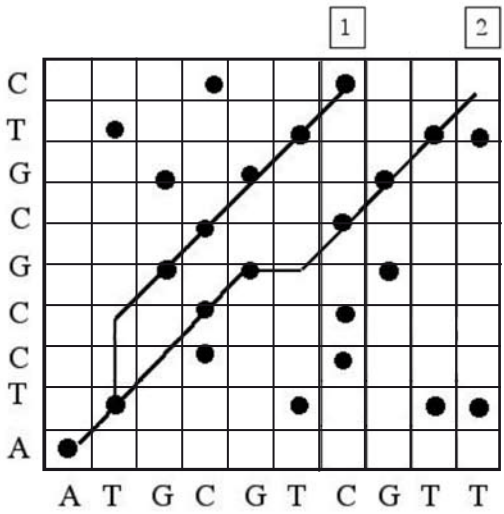
Como la secuencia 2 es una subsecuencia de la secuencia 1, hay una ruta continua a través de la diagonal de la matriz (ruta 1 en la figura 5) que corresponde al alineamiento:

```
ATGCGTCGTT
||||||||||
ATGCGTCGT
```

Pero se podrían crear otros alineamientos. En la figura 5 la ruta 2 que se desvía de la diagonal principal corresponde al alineamiento:

```
ATG - - - CGTCGTT
|||       |||
ATGCGTCGT
```

Figura 11. Una matriz de puntos para dos secuencias. En este ejemplo se muestran dos posibles "rutas" que corresponden a dos diferentes alineamientos posibles de las secuencias (Li y Graur, 1991)



En este caso la primera subsecuencia CGT de la secuencia 1 queda alineada con la segunda subsecuencia CGT de la secuencia 2. Esto requiere postular un hueco (en inglés *gap*, el término usado más habitualmente) en la secuencia 1 con una longitud de 3 bases. Además, esto es equivalente a formular la hipótesis de que en un momento dado ocurrió al menos un evento adicional: la inserción de las tres bases CGT en la secuencia 2, o bien, la supresión correspondiente en la secuencia 1 (estas dos alternativas, inserción o supresión, se engloban en el término *indel*). Ahora imaginemos alinear la secuencia 1 con una tercera secuencia de la siguiente manera:

```
Secuencia 1  ATGCGTCGTT
             | | | | | | |
Secuencia 3  ATCCGTCAT
```

En este caso las dos secuencias difieren en las posiciones tercera y octava; este emparejamiento imperfecto es llamado a menudo *mismatch*. La hipótesis implícita es que, como mínimo, ocurrieron dos sustituciones en una de las

secuencias desde que ambas surgieron de su último ancestro común. Por ejemplo, es posible que la secuencia ancestral fuera idéntica a la secuencia 1 y entonces la secuencia 3 habría sufrido al menos dos sustituciones. Otra posibilidad es que la secuencia ancestral fuera

secuencia ancestral hipotética ATACGTCAT

en cuyo caso tanto la secuencia 1 como la 3 habrían sufrido cada una al menos dos sustituciones, etcétera. Nótese que también es posible que desde la separación del último ancestro común ambas secuencias hayan sufrido múltiples sustituciones en las posiciones en las que difieren o incluso en aquellas en las que son idénticas.

En general, es posible alinear dos secuencias postulando alguna combinación de *indels* y sustituciones. Podemos calcular el costo de un alineamiento

$$C = s + ph$$

donde s es el número de sustituciones, h es la suma de las longitudes de los huecos y p es la penalidad por huecos que especifica el costo de un hueco relativo al de una sustitución. Por ejemplo, si los huecos son el doble de costosos que los emparejamientos imperfectos (mismatchs), entonces $p = 2$. La penalidad por huecos también implica que estamos haciendo una suposición acerca del proceso por el que divergieron las secuencias que tratamos de alinear; así, si creemos que las inserciones y deleciones son raras, p debería tener un valor alto.

La tarea del alineamiento de secuencias consiste en encontrar el alineamiento con el costo mínimo para una función o esquema de costos seleccionado. Existen varios algoritmos (Goad y Kanehisa, 1982; Waterman, 1984; Lipman y Pearson, 1985; Waterman *et al.*, 1991) que lidian con el alineamiento local de dos secuencias, que son particularmente útiles para trabajar con proteínas, ya que se trata de una aproximación más adecuada que el alineamiento global (Schuler, 2001).

Alineación de proteínas

Aunque el concepto de función de costo para la alineación de proteínas es similar al de nucleótidos, el modelo detrás del costo asignado a un *mismatch* entre aminoácidos es más complejo, pues hay que tomar en cuenta las dife-

rentes posibilidades a nivel de codones en las que un aminoácido puede ser sustituido por otro. Por ejemplo, una cisteína codificada por el triplete UGU puede ser remplazada por una tirosina mediante una única sustitución de nucleótidos (UAU), pero remplazarla con una metionina (AUG) requiere tres sustituciones de nucleótidos. Por lo tanto, alinear una cisteína con una tirosina es menos costoso que alinearla con una metionina. El costo del remplazo entre cualquier par de aminoácidos define una matriz de costos que puede usarse para evaluar cuantitativamente la calidad de un alineamiento.

Además de los costos implícitos en la sustitución de un triplete por otro, las matrices de sustitución de aminoácidos pueden reflejar otros criterios de asignación de costo como, por ejemplo, el grado de similitud fisicoquímica entre diferentes aminoácidos, o bien la frecuencia con la que determinada sustitución se ha observado empíricamente en bases de datos de alineamientos curados por expertos. Por ejemplo, las matrices BLOSUM (Henikoff y Heenikoff, 1992) se derivan de la base de datos Blocks, un conjunto de alineamientos sin huecos de regiones de familias de proteínas (Gibas y Jambeck, 2001). Las matrices PAM se derivan de un modelo de distancia de alineamientos de secuencias cercanamente emparentadas. Una "unidad" PAM es equivalente a un cambio promedio en el 1% de las posiciones consideradas en el alineamiento (Dayhoff *et al.*, 1978; Gibas y Jambeck, 2001).

Alineación de más de dos secuencias

Hasta ahora hemos enfocado la discusión en el alineamiento de pares de secuencias, pero la filogenia de dos entidades es trivial: entidad 1 – entidad 2. Así que si queremos reconstruir filogenias no triviales es necesario alinear más de dos secuencias. Teóricamente los algoritmos de programación dinámica (Eddy, 2004) usados para alinear pares de secuencias (Gupta *et al.*, 1995) podrían extenderse a cualquier número de ellas, pero los requerimientos de memoria de estos algoritmos crecen exponencialmente, por lo que la alineación de más de cuatro secuencias se vuelve prohibitiva (Thompson *et al.*, 1999). Para superar este problema se han desarrollado varias aproximaciones heurísticas y hoy en día existen numerosos programas que usan estrategias fundamentalmente diferentes (Thompson *et al.*, 1999). Tradicionalmente, el enfoque más popular ha sido el método de alineamiento progresivo (Feng y Doolittle, 1987). Programas de computadora basados en este método son MULTIALIGN (Barton y Sternberg, 1987), MULTAL (Taylor, 1988), PILE UP (Wisconsin Package) y CLUSTAL X (Thompson *et al.*, 1997), que es una

interfase gráfica para CLUSTAL W (Thompson *et al.*, 1994). Estos programas usan un algoritmo de alineamiento global (Needleman y Wunsch, 1970) para construir un alineamiento de la totalidad de la longitud de las secuencias.

En contraste, el programa PIMA (Smith y Smith, 1992) usa un algoritmo de programación dinámica local (Smith *et al.*, 1981) para alinear únicamente los motivos más conservados. Por su parte, el programa T-Coffee (Notredame *et al.*, 2000) se basa en la implementación de un algoritmo progresivo que puede corregir los alineamientos locales erróneos generados durante la etapa de comparaciones iniciales y se ha vuelto una de las herramientas con mejor desempeño cuando se trata de alinear secuencias poco conservadas entre sí gracias a la posibilidad de incorporar datos heterogéneos (evolutivos y estructurales) y de bloques conservados aislados a la información usada.

Recientemente han surgido numerosos algoritmos que ofrecen enfoques nuevos al problema del alineamiento múltiple, muchos de los cuales tienen en común aplicar estrategias iterativas para refinar y mejorar el alineamiento inicial. El programa DIALIGN (Morgenstein *et al.*, 1996) se basa en comparaciones segmento a segmento en lugar de residuo a residuo. El programa PRRP (Gotoh, 1996) optimiza un alineamiento progresivo global al subdividir iterativamente las secuencias en dos grupos que son subsecuentemente realineados. El programa SAGA (Notredame y Higgins, 1996) usa un algoritmo genético para seleccionar el alineamiento que optimiza una función objetivo.

Recientemente se ha comenzado a incorporar la tecnología de los modelos markovianos ocultos (HMM por sus siglas en inglés). Por ejemplo, el programa HMMT (Eddy, 1995) usa un método de recocimiento simulado para maximizar la probabilidad de que un HMM represente efectivamente a las secuencias por alinear. El programa MUSCLE (Edgar, 2004) usa una nueva función objetivo, llamada logaritmo de la expectativa, que estima rápidamente las distancias entre secuencias al basarse en la similitud de las subsecuencias en lugar de la similitud de los caracteres y hace un refinamiento del alineamiento al realinear todas las posiciones elegidas con criterios evolutivos. Todas estas características le confieren a MUSCLE una gran exactitud en sus alineamientos así como la capacidad de alinear en minutos conjuntos del orden de 5000 secuencias. Un inconveniente de MUSCLE es que sólo es capaz de alinear secuencias de aminoácidos, pero, en los casos en los que nos interese alinear secuencias codificantes, esta limitación se puede remediar mediante el uso de programas que transfieran la información de un alineamiento de proteínas a sus secuencias codificantes correspondientes, como, por ejemplo, el programa protal2dna (disponible como programa instalable en http://ftp.pasteur.

fr/pub/GenSoft/unix/alignment/protal2dna y como servidor accesible por la red mundial en http://bioweb.pasteur.fr/seqanal/interfaces/protal2dna.html). Quien se interese en comparaciones críticas de métodos de alineamiento puede consultar los artículos de Notredame (2000) y de Thompson et al., (1999). El sitio de red mundial del paquete de análisis filogenético PHYLIP (http://evolution.genetics.washington.edu/phylip/software.html) incluye una sección actualizada regularmente con referencias a programas de alineamiento.

Tras esta introducción a algunos aspectos fundamentales de la alineación de secuencias biológicas estamos listos para discutir cómo es que la información filogenética se puede ir difuminando con el paso del tiempo y qué podemos hacer para recuperarla.

El problema de las sustituciones múltiples

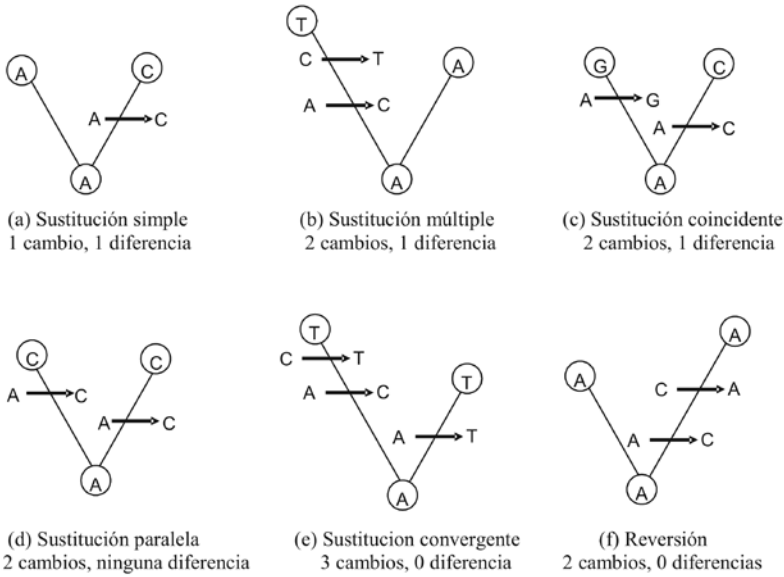
Cuando comparamos sitios homólogos en dos secuencias de ADN nuestra observación es simplemente si las secuencias, en esa posición, son iguales o no. Desde esa perspectiva el máximo número posible de diferencias entre dos secuencias en una posición dada es uno. Además, un sitio tiene uno de cuatro estados posibles (A, C, G o T si es ADN), lo que significa que si en un sitio dado ha ocurrido más de una sustitución desde el momento de la separación evolutiva de las secuencias que estamos comparando, perdemos todo registro de las sustituciones previas a la última. Por ejemplo, si en posiciones homólogas los seres humanos tienen una adenina (A) y los jitomates una guanina (G), sabemos que ha ocurrido al menos una sustitución desde que se separaron los linajes de los humanos y de los jitomates, pero es posible que hayan ocurrido muchas más sustituciones y de esas no podemos estar seguros (figura 12; Page y Holmes, 1998).

En el caso (a) de la figura 12 sólo ha ocurrido una sustitución en la historia de uno sólo de los dos linajes y eso se ve reflejado correctamente en el número de diferencias entre las secuencias. Nótese que el número de sustituciones que ha ocurrido efectivamente desde la separación de los linajes es pequeño, entonces es muy probable que la mayoría de las sustituciones sean sustituciones simples porque la probabilidad de que un sitio mute más de una vez es pequeña. En cambio, entre mayor sea el número de sustituciones, mayor es la probabilidad de que un mismo sitio haya sufrido múltiples sustituciones (lo que en inglés se denomina *multiple hits*).

En todos los casos excepto (a), el número de sustituciones que realmente ocurrió (sustituciones reales) es mayor al que estaríamos contando si simple-

mente comparáramos las dos secuencias descendientes. Los tres últimos casos pueden ser aún más engañosos, ya que las dos secuencias descendientes son idénticas y sin embargo esa similaridad no la heredaron directamente de su último ancestro común. Todos los métodos de inferencia filogenética dependen de que la similaridad entre secuencias u otros conjuntos de caracteres a comparar se deba principalmente a la homología y lo menos posible a la homoplasia (Page y Holmes, 1998).

Figura 12. Seis tipos de sustitución de nucleótidos.
En cada caso el nucleótido ancestral era *A* (Page y Holmes, 1998)



Si las mutaciones fijadas han sido relativamente comunes (o si la probabilidad de que una mutación se fije no es uniforme a lo largo de una secuencia), es posible que un sitio dado haya sufrido múltiples sustituciones que van borrando, a medida que pasa el tiempo y ocurren más sustituciones, la información o, si se prefiere, nuestro grado de certeza sobre las relaciones filogenéticas. Entre más tiempo pase, el simple número de diferencias entre secuencias es un indicador menos confiable del verdadero número de sustituciones que han ocurrido desde que las secuencias divergieron de su ancestro común (Page y Homes, 1998).

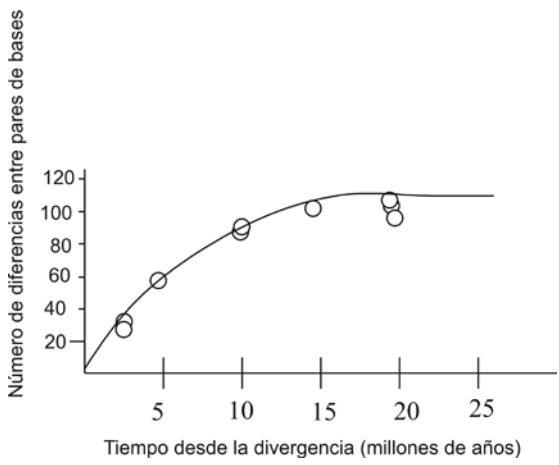
En el caso de la figura 13 los pares de secuencias que han divergido más recientemente muestran aproximadamente 17 sustituciones por cada millón de años. Esto nos haría esperar que los pares de especie que se separaron hace 20 millones de años mostraran alrededor de 340 diferencias. En realidad, entre ellas sólo se observan de 94 a 103 sustituciones.

Como se puede ver, a medida que se acumulan sustituciones entre dos secuencias éstas empiezan a saturarse, es decir, la mayor parte de los sitios que cambian ya han cambiado antes, por lo que las sustituciones más recientes tienen poco o ningún impacto en el número total de diferencias *observadas* entre las secuencias.

Modelos de sustitución

Dado que las distancias observadas pueden subestimar la verdadera magnitud del cambio evolutivo, se han desarrollado métodos de corrección de distancia para convertir las distancias observadas en medidas de distancia evolutiva verdadera. Muchos de ellos dependen de modelos que describen cómo se supone que están ocurriendo los procesos de evolución molecular. Estos modelos están relacionados entre sí, pero difieren en los aspectos del proceso de fijación de mutaciones que toman en cuenta. Por ejemplo, algunos

Figura 13. Número de sustituciones nucleotídicas entre pares de secuencias mitocondriales de bóvidos, contra el tiempo estimado de divergencia (Janecek *et al.*, 1996)



modelos pueden suponer que los diferentes nucleótidos no se encuentran en la misma proporción en las secuencias, o que las transiciones (mutaciones que cambian una purina por otra purina o una pirimidina por otra) ocurren con una frecuencia diferente que las transversiones (sustituciones de una purina por una pirimidina o viceversa), mientras que otros incluyen la posibilidad de que la tasa de sustitución varíe entre los diferentes sitios de las secuencias (Page y Holmes, 1998; Brinkamn y Leipe, 2001).

La idea fundamental detrás de los modelos de sustitución, ya sea que traten de secuencias de nucleótidos, de codones o de aminoácidos, es que ciertas sustituciones pueden ser más frecuentes que otras. En general, son más frecuentes las sustituciones entre bases bioquímicamente similares, entre codones que codifican para un mismo aminoácido y en aminoácidos similares. En el caso del ADN, los cuatro tipos de transiciones (de purina a purina o pirimidina a pirimidina, A→G, G→A, C→T, T→C) son habitualmente más frecuentes que los ocho tipos de transversión (de purina a pirimidina o el inverso, A→C, A→T, C→G, G→T y sus reversas), por lo que se plantea que las transiciones deben ser menos costosas e implicar una menor distancia que las transversiones.

Los esquemas de costos se pueden fijar *a priori* para que el método de inferencia filogenética calcule un costo exacto para cada sustitución postulada. Las matrices de costo fijo son equivalentes a las matrices de peso de estados de caracter (Brinkamn y Leipe, 2001; Morrone, 2000) y se aplican en la reconstrucción de filogenias por máxima parsimonia. Cuando se utilizan dichas matrices el método se llama más apropiadamente 'parsimonia ponderada' o 'parsimonia sopesada' (Wheeler, 1986; Goloboff, 1998). Para los métodos de reconstrucción basados en distancias o en máxima verosimilitud los costos se pueden derivar de matrices de tasas instantáneas que representan estimados de máxima verosimilitud de la probabilidad de que determinado tipo de sustitución ocurra. Una somera introducción al cálculo de las tasas instantáneas de cambio en el marco de la inferencia filogenética por máxima verosimilitud se encuentra en las presentaciones de Peter Foster (http://www.bioinf.org/molsys/data/idiots.pdf y http://www.bioinf.org/molsys/data/like.pdf).

Supongamos que la probabilidad de cualquier sustitución permanece constante a lo largo del tiempo y que la composición de bases nucleotídicas se encuentra en equilibrio. Entonces, la forma más simple de una matriz de sustituciones será:

$$P_T = \begin{matrix} P_{AA} & P_{AC} & P_{AG} & P_{AT} \\ P_{CA} & P_{CC} & P_{CG} & P_{CT} \\ P_{GA} & P_{GC} & P_{GG} & P_{GT} \\ P_{TA} & P_{TC} & P_{TG} & P_{TT} \end{matrix} \qquad \text{ecuación } A$$

donde P_{AC} es la probabilidad de que un sitio que al principio de un intervalo de tiempo *t* tenía el nucleótido A, se encontrará con el nucleótido C al final del intervalo. La mayoría de los modelos son simétricos, de manera que $P_{AC} = P_{CA}$. Los elementos que se encuentran en la diagonal corresponden al caso en que no ha habido (aparentemente) cambio en el sitio. Si asignamos los números 1-4 a los nucleótidos (A=1, C=2, etc.), entonces el valor de los elementos diagonales está dado por

$$P_{ij} = 1 - \sum_{j \neq i} P_{ij} \quad \text{ecuación } B$$

Dicho de otra forma, la probabilidad de observar A en un sitio dado en el tiempo 0 y nuevamente en el tiempo *t* es uno menos la probabilidad de observar la sustitución de A por C, G ó T. La composición de bases de la secuencia se puede representar mediante un vector

$$f = [f_A f_C f_G f_T] \qquad \text{ecuación } C$$

en el que f_A es la frecuencia en el equilibrio del nucleótido A, etc. En algunos modelos $f_A = f_C = f_G = f_T$, o sea que las cuatro bases están presentes en la misma proporción, mientras que en otros modelos pueden tener diferentes proporciones relativas (Page y Holmes, 1998).

Mediante la simple especificación de la probabilidad de una sustitución nucleotídica dada y de la frecuencia de bases esperada, se puede generar una enorme variedad de modelos de sustitución. El más simple y el primero en haber sido propuesto, es el de Jukes y Cantor (1969; modelo JC) y parte del supuesto de que las cuatro bases nucleotídicas tienen la misma frecuencia relativa y de que todas las sustituciones son igual de probables. Bajo este modelo la distancia entre dos secuencias está dada por

$$d = 3/4 \ ln(1 - 3/4 \ p) \quad \text{ecuación } D$$

donde *p* es la proporción de nucleótidos que son diferentes en las dos secuencias y ln es la función logaritmo natural. Usando las ecuaciones *A* y *C* podemos

representar el modelo Jukes-Cantor con la siguiente matriz de probabilidades de sustitución y el vector de composición de bases:

$$P_t = \begin{array}{c|cccc} & A & C & G & T \\ A & * & \alpha & \alpha & \alpha \\ C & \alpha & * & \alpha & \alpha \\ G & \alpha & \alpha & * & \alpha \\ T & \alpha & \alpha & \alpha & * \end{array} \qquad f = [1/4, 1/4, 1/4, 1/4]$$

donde α es la probabilidad de una sustitución y esa probabilidad es la misma para todas las sustituciones posibles (como hay doce sustituciones posibles, cada sustitución ocurre con probabilidad 1/12). Los valores de los elementos diagonales vienen dados por la ecuación *B* (Page y Holmes, 1998).

A pesar de que sólo una de cada tres sustituciones posibles es una transición, como ya señalé la observación empírica es que las transiciones son generalmente más frecuentes que las transversiones; el modelo de dos parámetros de Kimura (1980; modelo K2P) incorpora esta observación. En este modelo, si la tasa de transiciones por sitio es α y difiere de la tasa de transversiones (β), la tasa total de sustitución por sitio será (α+2β); por cada nucleótido hay tres cambios posibles, uno de los cuales es una transición y los otros dos, transversiones. En el modelo K2P el número de sustituciones nucleotídicas por sitio está dado por:

$$d = 1/2\ ln\ [1/(1-2P-Q)] + 1/4\ ln\ [1/(1-2Q)] \qquad \text{ecuación } E$$

donde *P* y *Q* son las diferencias proporcionales entre las dos secuencias comparadas debidas a transiciones y transversiones, respectivamente. Este modelo tiene la matriz de probabilidades de sustitución y el vector de frecuencias de bases:

$$P_t = \begin{array}{c|cccc} & A & C & G & T \\ A & * & \beta & \alpha & \beta \\ C & \beta & * & \beta & \alpha \\ G & \alpha & \beta & * & \beta \\ T & \beta & \alpha & \beta & * \end{array} \qquad f = [1/4, 1/4, 1/4, 1/4]$$

Obviamente, si $\alpha = \beta$ el modelo K2P se convierte en el modelo JC.

El modelo de Felsenstein (1981; modelo F81) fue el primero en permitir que las frecuencias relativas de los nucleótidos de las secuencias variaran, lo que estaría reflejando el hecho de que en determinadas secuencias, ciertos nucleó-

tidos son más frecuentes que otros y por lo tanto ciertas sustituciones tienen mayor probabilidad que otras. El modelo F81 tiene la siguiente forma

$$P_t = \begin{array}{c|cccc} & A & C & G & T \\ A & * & \Pi_C\alpha & \Pi_G\alpha & \Pi_T\alpha \\ C & \Pi_A\alpha & * & \Pi_G\alpha & \Pi_T\alpha \\ G & \Pi_A\alpha & \Pi_C\alpha & * & \Pi_T\alpha \\ T & \Pi_A\alpha & \Pi_C\alpha & \Pi_G\alpha & * \end{array} \qquad f = [\Pi_A, \Pi_C, \Pi_G, \Pi_T]$$

donde Π_i es la frecuencia de la *i*-ésima base promediada sobre las secuencias comparadas. Nótese que si $\Pi_A = \Pi_C = \Pi_G = \Pi_T$ entonces el modelo F81 es igual al JC.

Modelos aún más refinados aparecieron posteriormente, incluyendo el modelo de Hasegawa, Kishino y Yano (1985) y el de Kishino y Hasegawa (1989), que esencialmente conjuntaban las aportaciones de los modelos K2P y F81 al generar un modelo que permitía tanto un sesgo en la proporción de transiciones y transversiones como frecuencias desiguales de nucleótidos. Esta tendencia culminó con la formulación del modelo general reversible en el tiempo (*GTR* por sus siglas en inglés) que no sólo permite frecuencias desiguales de nucleótidos, sino también que cada una de las seis posibles rutas de sustitución (A↔C, A↔G, A↔T, C↔G, C↔T y G↔T) sea recorrida a una tasa diferente (Rodríguez *et al*., 1990).

En un intento de hacer una descripción lo más fiel posible de los procesos que subyacen a la divergencia de secuencias, los modelos han continuado refinándose. Por ejemplo, se ha añadido la variación en las tasas de sustitución en los diferentes sitios de las secuencias (Churchill *et al*., 1992; Yang, 1993, 1994). Uno de los métodos más comúnmente aplicados para ello es el gama discreto, que presupone que las tasas relativas se distribuyen de acuerdo a una distribución gama. Valores grandes del parámetro α de esta distribución corresponden a una pequeña varianza en las tasas relativas (las tasas de todos los sitios son casi las mismas) mientras que valores pequeños llevan a una distribución sesgada de las tasas relativas (la mayoría de las tasas son pequeñas pero algunos sitios tienen tasas muy elevadas de sustitución; Swofford *et al*., 1996; Yang, 1996; Huelsenbeck y Crandall, 1997; Lewis, 1998, 2001b).

Algunos programas permiten la modelación de la heterogeneidad de las tasas entre los sitios, por ejemplo sitios totalmente invariables (Swofford, 2000); modelos de tasas específicas a sitios, en los que existe una tasa de sustitución diferente para las diferentes posiciones dentro de genes, como introns y exones, así como para diferentes regiones de una proteína o posiciones dentro

de los codones (Swofford, 2000); modelos markovianos ocultos que permiten un grado de correlación entre las tasas de sitios que son adyacentes (Felsenstein, 1999) y modelos gama discretos en los que el parámetro α puede variar entre subconjuntos de los sitios (Yang, 1997). También es posible combinar un modelos de sitios invariables con un modelo gama discreto (Swofford, 2000; Lewis 2001b).

Modelos que incorporan falta de independencia y selección natural

Otros modelos, conocidos como modelos de codones y modelos de estructura secundaria (Lewis, 2001b) permiten que la distribución de tasas de sustitución refleje la dependencia entre los sitios nucleotídicos de un mismo codón y entre los sitios que se aparean en las regiones de tallo de un gen de rARN, respectivamente. Esto contrasta con los métodos basados en parsimonia y los métodos tradicionales basados en máxima verosimilitud, en los que se presupone que todos los sitios son evolutivamente independientes. Por ejemplo, en genes codificantes para proteínas los tres sitios nucleotídicos que forman un codón no pueden evolucionar independientemente unos de otros si existe selección para el aminoácido correspondiente (Lewis, 2001b). Muse y Gaut (1994) y Goldman y Yang (1994) desarrollaron independientemente modelos que toman en cuenta este aspecto. Para mantener un número razonable de parámetros introducen algunas restricciones, por ejemplo, no se permite más de una sustitución en un instante dado (la tasa del cambio AAA $\rightarrow$ AGC será cero), pero se permite que ocurran dos cambios consecutivos en períodos infinitesimales de tiempo.

El modelo de Muse y Gaut asigna tasas a todos los demás cambios posibles (que requieren más de dos mutaciones puntuales) con base en dos parámetros que representan la tasa de sustituciones sinónimas y no-sinónimas respectivamente (Muse y Gaut, 1994; Lewis, 2001b). De esta forma, estimando sólo cinco parámetros (los dos parámetros de tasa mencionados y tres parámetros de frecuencia de nucleótidos) se puede especificar toda la matriz de $61 * 61$ probabilidades de transición. El modelo de Hashegawa, Kishino y Yano (1985) que se mencionó anteriormente también requiere cinco parámetros, pero en su caso los dos parámetros de tasas representan las tasas de transiciones y transversiones.

El modelo de Goldman y Yang (1994) va un paso más allá, al permitir tanto un sesgo de transiciones/transversiones como uno de sustituciones

sinónimas/no-sinónimas. Además, este modelo permite que las diferencias bioquímicas entre los aminoácidos especificados por cada codón jueguen un papel en la determinación de la tasa a la que ocurren los cambios entre aminoácidos. Por ejemplo, el cambio de un residuo hidrofóbico por otro puede ocurrir a una tasa más alta que el cambio de uno hidrofóbico por uno hidrofílico (Lewis, 2001b). Los modelos de codones están implementados en los programas computacionales HYPHY (Muse y Kosakovsky-Pond, 2000) y PAML (Yang, 1997), pero en esos programas están enfocados sobre todo a cálculos posteriores a la búsqueda de árboles filogenéticos. En cambio, el programa MrBayes (Huelsenbeck y Ronquist, 2001) utiliza los modelos de codones en la búsqueda de árboles filogenéticos.

También se ha intentado añadir realismo a los modelos de sustitución para describir la dependencia evolutiva de sitios nucleotídicos que están frente a frente en las regiones de tallo de los RNAs con estructura secundaria (Lewis, 2001b), como los desarrollados independientemente por Tillier y Collins (1995) y Muse (1995). Recuérdese que la selección natural favorece las mutaciones compensatorias en estos sitios para preservar la estabilidad de las regiones de tallo y la estructura secundaria necesaria para la actividad catalítica de estas moléculas de ARN.

Modelos de cambio en aminoácidos

Al igual que con las secuencias de nucleótidos, la divergencia entre secuencias de aminoácidos puede modelarse recurriendo a una matriz de sustituciones (que denotaremos como M) que describa las probabilidades, para un tiempo dado, de las mutaciones (o más apropiadamente, sustituciones o reemplazos) del aminoácido i al aminoácido j

$$Pr\{\text{aminoácido } i \rightarrow \text{aminoácido } j\} = M_{ij}$$

Esto corresponde a un modelo en el que los aminoácidos mutan aleatoriamente e independientemente unos de otros pero de acuerdo con probabilidades predefinidas asociadas a cada aminoácido; esas son propiedades de un modelo markoviano de evolución y aunque es muy simple, es uno de los mejores. Se pueden construir matrices que modelen las propiedades de los aminoácidos tales como tamaño, carga, hidrofobicidad, etc.; sin embargo, la modelación de dependencias que describan cómo se relacionan las características de un aminoácido con las características de sus vecinos no es posible

con este mecanismo (aunque se pueden incluir derivaciones del modelo markoviano que permitan dependencias -es interesante notar que ningún esquema de costos de sustitución actualmente en uso considera la posibilidad de aumentar la duración de su "memoria", es decir calcular el valor de la alineación de oligómeros en lugar de bases o aminoácidos individuales, lo que equivaldría a usar modelos markovianos de orden superior). Por otra parte, las diferentes frecuencias con las que los aminoácidos aparecen en la naturaleza se pueden incorporar al modelo, aquí las denotaremos con *fi*. El modelo es generalmente simétrico, es decir, que la probabilidad de que el aminoácido *i* mute a *j* es la misma que comenzar con *j* y terminar con *i*.

En contraste con los modelos de sustitución de nucleótidos y de codones, los de remplazo de aminoácidos se han basado en un enfoque empírico. Margaret Dayhoff y sus colaboradores elaboraron un modelo de evolución de proteínas que desembocó en la creación de un conjunto de matrices de sustitución ampliamente utilizado (Dayhoff *et al*., 1978). En esta aproximación, las tasas de remplazo son derivadas de alineaciones de secuencias de proteínas que son por lo menos 85% idénticas; esta restricción asegura que la probabilidad de que una sustitución observada sea consecuencia de un conjunto de mutaciones sucesivas sea baja. Las matrices Dayhoff se utilizan frecuentemente en la exploración de bases de datos basada en el alineamiento de pares de secuencias, así como en el alineamiento de múltiples secuencias, pero también son de utilidad en aplicaciones filogenéticas.

En nuestra definición, la matriz *M* implica cierta cantidad de mutación (medida en unidades PAM -por las siglas en inglés de *point accepted mutations*, es decir, mutaciones de una posición aceptadas o fijadas o sustituciones realmente observables al comparar secuencias). Una matriz de mutación PAM1 se refiere a una cantidad de mutaciones que hará que cambien, en promedio, el 1% de los aminoácidos. Los elementos diagonales de *M* son las probabilidades de que un aminoácido dado no cambie y $(1\text{-}M_{ii})$ es por lo tanto la probabilidad de mutar de *i* a cualquier otro aminoácido. Si tenemos un vector de probabilidades *p*, el producto *Mp* nos da las frecuencias esperadas después de un período evolutivo equivalente a una unidad PAM1. Nótese que el tiempo cronológico no depende linealmente de las distancias PAM y que las tasas de evolución pueden ser muy variables para diferentes especies y diferentes proteínas.

Dayhoff y colaboradores (1978) presentaron un método para estimar la matriz *M* a partir de la observación de 1572 mutaciones aceptadas entre 34 superfamilias de secuencias cercanamente emparentadas. Una matriz

Dayhoff se calcula a partir de la matriz de mutaciones PAM-250. Las celdas de la matriz Dayhoff se relacionan con la matriz PAM-250 mediante la fórmula:

$$D_{ij} = 10log_{10} [(M^{250})_{ij} / f_i]$$

Más recientemente, Jones *et al.* (1992) y Gonnett *et al.* (1992) usaron una metodología similar a la de Dayhoff pero con bases de datos modernas. El modelo de Jones y colaboradores ha sido implantado en análisis filogenéticos con cierto éxito. Jones *et al.* (1994) también han calculado una matriz de remplazo de aminoácidos específicamente para proteínas transmembranales, que tiene valores notablemente diferentes de los de las matrices Dayhoff, las cuales están sesgadas hacia proteínas globulares solubles en agua, sobre representadas en la muestra original de 34 superfamilias.

Adachi y Hasegawa (1995, 1996) pusieron a punto un modelo general markoviano reversible de sustitución de aminoácidos que usa una matriz derivada de los reemplazos inferidos entre proteínas mitocondriales de veinte especies de vertebrados. Estos autores han mostrado que este modelo tiene mejor desempeño que otros al tratar con filogenias de proteínas mitocondriales.

Henikoff y Henikoff (1992) usaron una aproximación diferente: aprovecharon su base de datos de alineamientos locales sin huecos de proteínas lejanamente emparentadas para derivar la serie de matrices BLOSUM. Las matrices de esta serie se identifican con un número (por ejemplo, BLOSUM50) que se refiere al porcentaje mínimo de identidad de los bloques de aminoácidos alineados usados para construir la matriz. Estas matrices se calculan directamente sin extrapolaciones y son análogas a matrices de probabilidades de transición *P(T)* para diferentes valores de *T*, estimadas sin referencia a ninguna matriz de tasas *Q*. A menudo, las matrices BLOSUM tienen un mejor desempeño que las matrices PAM en búsquedas de similitud local pero no han sido usadas muy ampliamente en filogenias.

Nei (1987) propuso un modelo no-empírico simple de sustitución de aminoácidos que supone que las frecuencias de sustituciones siguen una distribución Poisson. Este modelo da estimaciones precisas del número de sustituciones de aminoácidos cuando se comparan especies cercanamente emparentadas.

Algunos programas de análisis filogenético, notablemente PHYLIP incluyen la posibilidad de calcular matrices de distancia para aminoácidos con el método de Kimura (1983). Se trata de una fórmula para aproximar la distancia

PAM midiendo la fracción de aminoácidos *p* que difiere entre dos secuencias y computar la distancia como

$$D = l_n \ (1 - p - 0.2\ p^2)$$

Es un método muy rápido para obtener el número de sustituciones por proteína pero tiene algunas limitaciones, ya que no toma en cuenta qué aminoácidos difieren o hacia qué aminoácidos cambian, de manera que se pierde información. Las unidades de la medida de distancia son la fracción de aminoácidos que difieren, como en el caso de la distancia PAM. Si la fracción de aminoácidos que difiere es mayor a 0.8541, la distancia se vuelve infinita.

Modelos que permiten variación de las tasas entre diferentes posiciones de las secuencias

Los primeros modelos de evolución de secuencias suponían que todos los sitios variables evolucionaban a la misma tasa (Penny *et al*., 2001), pero pronto se introdujeron dos aproximaciones que permitían flexibilizar esta suposición. Una de ellas consiste en ajustar una distribución de probabilidad para las tasas (Uzzell y Corbin, 1971); la otra fue el llamado modelo de covarión (Fitch y Markovitz, 1970; Fitch, 1971). Bajo el modelo de Uzzell y Corbin de tasas a través de sitios (*RAS* por sus siglas en inglés), cada sitio tiene una tasa característica (o intrínseca) que se mantiene durante todo el período de tiempo abarcado por la divergencia de las secuencias que se están estudiando. Los sitios difieren en esta tasa intrínseca y la variación se puede modelar con una distribución de probabilidad. Para un análisis detallado ver Chang (1996). Un avance significativo fue la introducción por Yang (1993, 1994, 1996) de un modelo que permitía tasas de sustitución variables entre sitios dentro de un marco de máxima verosimilitud. Yang mostró que añadir un solo parámetro -la forma de una supuesta distribución gama para las tasas- aumentaba la verosimilitud de los análisis por muchos órdenes de magnitud (Galtier, 2001). Presumiblemente esto se debe a que las restricciones selectivas varían entre los sitios. Tomar en cuenta este efecto mejoraría enormemente nuestra representación de la evolución molecular. Además de la distribución gama, se han usado otras distribuciones teóricas (por ejemplo, Waddell *et al*., 1997) y distribuciones medidas empíricamente (Van de Peer *et al*., 1996).

Sin embargo, un problema de los modelos RAS es que requieren estrictamente que cada sitio mantenga su tasa durante toda la evolución y que no hay una explicación bioquímica de porqué un sitio debería mantener su tasa intrínseca por largos períodos evolutivos. Cuando se consideran largos períodos de tiempo, los sitios que son críticos para la función de una macromolécula pueden cambiar, haciendo que la tasa evolutiva de un sitio dado cambie a lo largo de la filogenia. Esta idea se introdujo hace más de treinta años por Fitch y Markowitz (1970) y Fitch (1970) quienes lo llamaron proceso de covarión. Esta terminología viene de la idea de que la tasa de un sitio puede ser modificada por una sustitución que ocurra en un sitio diferente de la molécula, con el cual, por lo tanto, covariaría. El modelo de covarión postula que aunque algunos sitios son críticos para la función y no pueden cambiar nunca, la mayor parte de los sitios puede alternar entre evolucionar libremente en algunos taxa, o ramas de la filogenia, y estar fijados en otros, alternancia que sería resultado de cambios leves en la estructura secundaria y terciaria de las macromoléculas. En la mayor parte de los programas computacionales de análisis filogenético (PAUP*, MEGA2, Phylip, MrBayes, HyPhy, PAML, etc.) es posible usar modelos que incorporen variación de tasas entre sitios de tipo RAS. La incorporación de heterogeneidad de tasas con modelos de tipo covarión no se ha implementado ampliamente aunque se encuentra disponible en la versión 3 de MrBayes. Actualmente es posible imaginar extensiones a los modelos de covarión (Penny *et al.,* 2001) como incorporar una categoría adicional de sitios que no variarían en absoluto a lo largo de la historia evolutiva u otros que nunca estarían restringidos, que podrían mejorar la detección de sitios que evolucionan lentamente, lo que sería de gran utilidad para estudiar divergencias muy antiguas.

Elección del modelo

Para los análisis filogenéticos basados en distancias, como los de máxima verosimilitud, el investigador debe escoger un modelo para describir cómo evolucionaron las secuencias que está analizando (aún si esto significa simplemente escoger una familia de modelos y dejar que el programa de estimación de filogenias infiera los parámetros [Holder y Lewis, 2003]). Entre más complejo sea un modelo, mejor se ajustará a los datos, pero también aumentará más la varianza de los parámetros estimados. Las estrategias de selección de modelos (Posada y Crandall, 2001; Goldman y Wheelan , 2000; Ota *et al*., 2000; Suchard *et al*., 2001; Abascal *et al*., 2005) tratan de encontrar el nivel de complejidad

apropiado con base en los datos disponibles. Una complejidad excesiva del modelo puede llevar a que los datos sean intratables computacionalmente (Holder y Lewis, 2003) de manera que las preocupaciones pragmáticas pueden pesar más que la corrección estadística, lo que hace atractiva la velocidad de los métodos de *neighbor joining* y de parsimonia.

Existen programas para seleccionar el modelo más adecuado para describir una alineación, tanto de nucleótidos (Posada y Crandall, 1998; Nylander, 2004) como de aminoácidos (Abascal *et al*., 2005). Asimismo, existen pruebas para determinar si para un juego de datos dado es más adecuado un modelo tipo RAS o uno tipo covarión (Tuffley y Steel, 1997; Lockhart *et al*., 1998, 2000).

Secuencias no alineables y métodos que no requieren alineación

Cuando las secuencias de diferentes taxa varían en longitud, la correspondencia u homología entre las diferentes posiciones puede ser difícil de identificar sin ambigüedad. Cuando se estudian divergencias profundas, la proporción de sitios afectados por esta ambigüedad puede ser alto, a veces de más del 40% (Lee, 2001; Wheeler, 1995; Lutzoni *et al*., 2000; Gatesy *et al*., 1993; Janke *et al*., 2001; Stanhope *et al*., 1998). La mayor parte de los estudios filogenéticos adoptan la estrategia de descartar, antes de los análisis, los sitios que no se pueden alinear inequívocamente entre todos los taxa (Swofford *et al*., 1996), bajo el supuesto de que contienen poca información evolutiva. Por ejemplo, las terceras posiciones de los codones son habitualmente descartadas bajo el pretexto de que evolucionan demasiado rápido, pero pueden contener una mejor señal filogenética que las posiciones primera y segunda (Lee, 2001; Sennblad y Bremer, 2000). En general, al tratarse de las regiones que evolucionan más rápidamente de los genomas, los sitios de alineamiento ambiguo pueden contribuir a resolver divergencias temporalmente muy cercanas y a estimar mejor las tasas de evolución molecular.

Los primeros métodos que trataron de rescatar los sitios ambiguos trataban de identificar todos los alineamientos plausibles mediante programas que contemplaran todo un rango de parámetros relevantes (transiciones/transversiones, apertura de huecos, extensión de huecos, etc.; Lee, 2001). Los alineamientos múltiples generados con diferentes combinaciones de valores de los parámetros se analizan separadamente y sólo se aceptan las relaciones filogenéticas que aparecen en todos (o la mayoría) de los alineamientos. La mayor desventaja es la necesidad de numerosos análisis que toman mucho tiempo.

En el método de elisión también se genera un rango de alineamientos plausibles pero en lugar de analizarlo separadamente, se concatenan en una única matriz grande y se evalúan en un solo análisis (Lee, 2001; Wheeler, 1995; Thornton y DeSalle, 2000). Habrá regiones que a pesar de ser ambiguas se alinearán consistentemente para determinados grupos de taxa y aparecerán en todas las combinaciones de parámetros por lo que tendrán el mismo peso que las regiones completamente libres de ambigüedad; mientras que las regiones cuyo alineamiento varíe entre juegos de parámetros tienen un menor peso, pero aún en esas regiones se rescatan bloques consistentes.

Una crítica posible al método de elisión es que no toma en serio el concepto de homología, pues cada base debe tener una historia individual pero no es tratada como si la tuviera (Lutzoni *et al*., 2000). Es cierto que la homología es una relación de todo o nada, pero la precisión de la inferencia de homología (es decir, el alineamiento) no es absoluto sino probabilístico (Lee, 2000; Mitchinson, 1999) por lo que parece razonable darle un peso diferente a cada una de esas inferencias. Al darle menor peso a los alineamientos que se encuentran sólo en regiones restringidas del espacio de parámetros, el método de elisión hace la suposición razonable de que tales alineamientos son menos probables. Por otro lado, existe un problema metodológico que tiene que resolverse: la implícita duplicación de matrices (y de caracteres) en el método de elisión plantea problemas para la estimación de la robustez de un árbol, porque las aproximaciones más populares (el *bootstrap* no paramétrico y las pruebas Kishino-Hasegawa y Shimoidara-Hasegawa, [Goldman *et al*., 2000]) son métodos de remuestreo que suponen independencia de caracteres.

Recientemente Wheeler (1999) y Lutzoni *et al*. (2000) desarrollaron independientemente un método que trata la totalidad de una región de alineamiento ambiguo como un único caracter homólogo cuyos diferentes estados de caracter posibles corresponden a cada distinta variante de la secuencia (Lee, 2001). Este método, llamado alineamiento a nivel de fragmento, considera todas las alineaciones posibles, pero incluye en el análisis final sólo aquella que es más consistente con el resto de la evidencia presentada por los caracteres, de manera que no plantea interrogantes respecto al concepto de homología subyacente. Una dificultad práctica, en cambio, es que si la región ambigua es muy variable, el número de estados de caracter para esa región puede volverse intratable para programas convencionales de análisis filogenético, como PAUP* (Swofford, 2000), aunque tales dificultades pueden evitarse en POY (Giribit, 2001; Gladstein y Wheeler, 1996), un programa diseñado para este tipo de análisis. Una limitante quizá más significativa es

que la matriz de alineación a nivel de fragmento, a diferencia de la de elisión, es muy difícil de analizar por máxima verosimilitud, aunque el desarrollo de modelos de máxima verosimilitud para datos no moleculares (Lewis, 2001b) podría conducir a métodos estadísticos explícitos que sirvieran para este tipo de alineamientos.

Los fundamentos del análisis comparativo de secuencias que hemos discutido hasta ahora contienen un sesgo relativo al trato que se da a las moléculas biológicas como secuencias lineales de unidades discretas similares a las representaciones lingüísticas, pasando por alto, en cierta medida, su naturaleza como estructuras tridimensionales y el carácter dinámico de la evolución molecular (Vinga y Almeida, 2003). La recombinación genética y en particular el rearreglo (*shuffling*) a nivel genómico son una fuente de conflicto potencial para la comparación de secuencias mediante alineamientos, pues los rearreglos rompen la contigüidad que se presupone está conservada entre segmentos homólogos. Además, actualmente no existen métodos estadísticos para estudiar la evolución de genomas completos (Li *et al*., 2001) que permitan modelar los procesos específicos de la evolución en ese nivel de organización (rearreglos, duplicaciones, etc.). Con el fin de superar esas limitaciones, recientemente se ha formulado una variedad de métricos de similaridad entre secuencias que no dependen de comparaciones posición por posición y que han llevado a implementar algoritmos que pueden ser usados en la reconstrucción filogenética a partir de genomas completos, así como en sistemas rápidos de clasificación y en el análisis filogenético basado en secuencias relativamente cortas pero muy difíciles de alinear. Muchos de estos métodos son robustos ante la variación extrema del tamaño de las secuencias comparadas y algunos son robustos ante la repetición interna de motivos, dos problemas que los métodos basados en alineamientos no han resuelto del todo.

Los métodos de comparación independientes del alineamiento se pueden agrupar en dos grandes categorías: los métodos que analizan la frecuencia de "palabras" (sub-segmentos de longitud fija dentro del conjunto de datos; en general, si la longitud dada es de L símbolos, los sub-segmentos se denominan L-palabras o L-tuplos) y los que no involucran el conteo de estos segmentos de longitud fija.

Los métodos basados en la frecuencia de palabras comienzan haciendo un mapeo de las secuencias a vectores (listas) definidos por el conteo de cada L-tuplo. La lógica es que secuencias similares compartirán en cierto grado su composición de palabras. De cierta forma esto es una extensión del

uso de la diferencia en contenido de GC como una medida de la disimilitud de secuencias. Hay que señalar que los métodos basados en frecuencias de palabras, a pesar de ser libres de alineamiento, aún son dependientes de la longitud pues las comparaciones se hacen para una longitud de palabras fija. Sin embargo, algunos utilizan la información derivada de todas las posibles longitudes de palabras para poder lograr completa independencia de la contigüidad de los segmentos conservados. Entre los métodos basados en frecuencias de palabras se pueden destacar la distancia euclidiana propuesta por Blaisdell (1986) quien fue el primero en mostrar la posibilidad de medir la disimilitud entre secuencias demasiado diferentes como para poder ser alineadas, aunque todavía se puedan reconocer como emparentadas. Es notable que con este método pionero se pudieran corroborar relaciones filogenéticas obtenidas con métodos convencionales. El método de distancias euclidianas se ha generalizado para volverse de distancias euclidianas ponderadas, que ha sido incorporado en el programa de cómputo STACK (Christoffels *et al*., 2001). Otros métodos basados en frecuencias de palabras se basan en la estructura de correlación (a grandes rasgos, el concepto describe cómo la presencia de una subsecuencia afecta la probabilidad de la presencia de otra subsecuencia), un enfoque utilizado para clasificar proteínas basándose en sus frecuencias de dipéptidos (Petrilli, 1993). Otro métrico interesante se ha propuesto recientemente (Stuart *et al*., 2002a y b) en el que la distancia entre dos secuencias se basa en el ángulo entre los vectores de conteo de los *L*-tuplos. Este métrico tiene la propiedad interesante de no ser sensible a las repeticiones, pues si la secuencia X es comparada con su repetición XX, sus vectores de conteo tendrán la misma dirección en el espacio y el ángulo entre ellos será cero. Esto se traduce en un filtrado automático de repeticiones, lo que permite distinguir a las secuencias sólo con base en su composición de *L*-tuplos (Vinga y Almeida, 2003).

Además, existen al menos dos métodos que no dependen de longitudes fijas de palabras, lo que constituye una independencia absoluta del supuesto de conservación de contigüidad entre caracteres. El primero de estos métodos (llamado *mapas universales de secuencia*) se basa en la propuesta de utilizar funciones iterativas que aparecen frecuentemente en la teoría del caos para representar secuencias biológicas (Jeffrey, 1990; Almeida y Vinga, 2002). El segundo enfoque se basa en usar la compresión informacional de las secuencias como una medida de su complejidad. En este contexto, la compresión informacional se refiere a los procesos que permiten que un archivo de cómputo ocupe menos espacio de memoria mediante la eliminación de la redundancia

sin perder la parte esencial de la información que contiene. Los lectores tal vez estén familiarizados con la compresión informacional pues es la idea detrás de programas como StuffIt y de formatos como el MP3. El grado en el que la compresión conjunta de pares de secuencias es más efectiva que la compresión independiente se puede usar como medida de la similaridad. Concretamente, Li *et al*. (2001) usan el programa GenCompress (Chen *et al*., 1999) para comprimir los archivos que contienen cada par de las secuencias que se quieren comparar. La proporción en que se pueden comprimir los archivos con las secuencias yuxtapuestas es una medida de la similitud entre las secuencias. Todos los métodos de comparación libres de alineamiento expuestos aquí han sido implementados por Almeida y Vinga (2002) en programas en el lenguaje MATLAB y están disponibles en http://www.bioinformatics.musc.edu/resources.html. Sería interesante que se lograra su implementación en lenguajes de más amplio acceso. Es importante subrayar que los métodos de comparación de secuencias independientes de alineamientos, a pesar de ser muy poderosos en los casos en los que no se puede resolver la contigüidad de los caracteres, conllevan un costo que quizá sea muy alto si nuestro interés en la filogenia consiste en estudiar la evolución de los caracteres con los que se infiere dicha filogenia. En efecto, los métodos libres de alineamiento desechan la información relativa a la comparación caracter por caracter y para fines prácticos tratan cada secuencia como si tuviera un solo caracter, por lo que se imposibilitan estudios que busquen elucidar el orden de cambio de los caracteres contenidos en las secuencias comparadas. Sin embargo, queda la posibilidad de usar los métodos libres de alineamiento para reconstruir una filogenia con el juego de secuencias A y suponer que dicha filogenia es válida para el juego de caracteres B cuya evolución se estudiará.

Métodos de estimación de filogenias

Los métodos de inferencia de árboles filogenéticos se pueden clasificar en aquellos basados en distancias y los basados en caracteres. Los primeros calculan las distancias entre cada par de secuencias de acuerdo a alguna medida predeterminada, generalmente la cantidad de disimilitud entre secuencias, y desechan los datos reales, usando únicamente las distancias fijadas para derivar los árboles. Los segundos derivan los árboles que optimizan la distribución de los patrones de datos reales para cada caracter; las distancias pareadas no son fijadas sino determinadas por la topología del árbol (pero de todas formas, dado un árbol, existe una distancia entre cada par de secuencias), y por ello

se ha propuesto que sería más correcto que los métodos de distancias y los métodos de caracteres se llamaran, respectivamente, métodos de construcción de árboles y métodos de búsqueda de árboles.

Métodos basados en distancias

Un método de distancias podría reconstruir el árbol verdadero (el que representara la serie de eventos de divergencia entre las unidades taxonómicas estudiadas en el orden en que ocurrió) si todos los eventos estuvieran adecuadamente representados en las secuencias (Swofford *et al*., 1996; Brinkman y Leipe, 2001). Sin embargo, como vimos anteriormente, al saturarse las secuencias el grado de divergencia alcanza un límite máximo, por lo que muchos métodos basados en distancias buscan corregir esto mediante el uso de una matriz de tasas que supone que una proporción de las identidades observadas debe de ser consecuencia de mutaciones múltiples, y que esa proporción aumenta junto con el grado de divergencia total. Estas correcciones han sido incorporadas por programas de estimación de filogenias como MEGA 2.1 (Kumar *et al*., 2001) que incorpora distancias corregidas como la de Tamura-Nei para comparar secuencias de nucleótidos, o las distancias corregidas de Poisson o gama al comparar proteínas. La corrección de distancia de Poisson corrige para sustituciones múltiples en un sitio dado pero presupone igualdad en las tasas de sustitución entre sitios e igualdad en las frecuencias relativas de los aminoácidos. La distancia gama representa una mejoría respecto a la corrección de Poisson al tomar en cuenta la desigualdad en las tasas de substitución entre sitios; sin embargo, para que sea de utilidad, requiere que se estime el parámetro α que describe la forma de la distribución gama. Valores habituales de este parámetro son α =2.25 cuando se emplean las matrices Dayhoff para comparar las secuencias, o bien α =0.65 cuando se calcula la distancia Grishin, pero estos valores sólo se dan aquí a título indicativo (ver Nei y Kumar, 2000). El programa PAUP 4.0 (Swofford, 2000) incluye la posibilidad de usar un modelo estimado por máxima verosimilitud a partir de las propias secuencias que se comparan.

Los métodos de distancia son menos demandantes en poder de cálculo y mucho más rápidos que los basados en máxima verosimilitud pero apelan a los mismos modelos de evolución molecular, siendo probablemente ésta su principal ventaja. La desventaja es que, para fines prácticos, los datos reales son desechados (Brinkman y Leipe, 2001b; Lewis, 2001; Morrone, 2001) es

decir, la información contenida en la comparación caracter por caracter se colapsa en una única medida de distancia entre cada par de secuencias.

* Método de agrupamiento por pares no ponderados con media aritmética (UPGMA)

Es un algoritmo de agrupamiento que une ramas en un árbol basándose en el criterio de maximizar la similitud entre pares y entre promedios de pares ya unidos. En sentido estricto, no es un método de distancia evolutiva (Li, 1997). UPGMA puede generar una topología acertada con las longitudes de ramas verdaderas sólo si la divergencia entre secuencias se apega a un reloj molecular (Swofford *et al*., 1996), condición que raramente se cumple en la práctica.

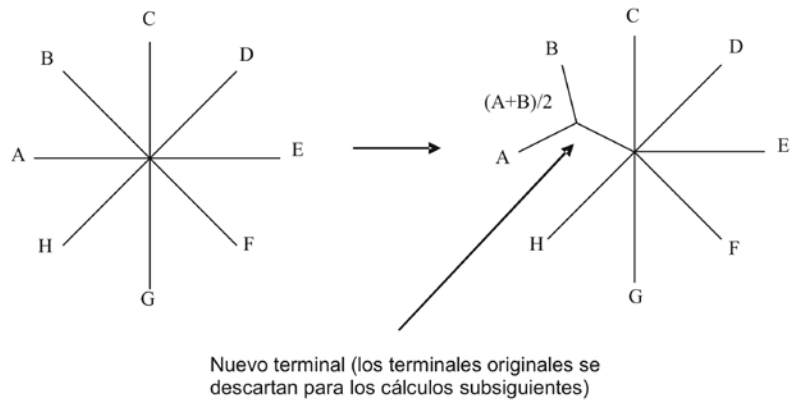
* Neighbor joining

Se aplica habitualmente cuando se construyen árboles de distancia, independientemente del criterio de optimización. El árbol completamente resuelto se 'descompone' a partir de un árbol 'estrella' (completamente no resuelto) por el procedimiento de insertar ramas entre el par de vecinos más cercanos (o más aislados de los demás) y el resto de los nodos terminales del árbol (figura 14). El par de vecinos más cercanos queda entonces consolidado dando por resultado un nuevo árbol 'estrella' y el proceso se repite hasta obtener un árbol resuelto. Es un método comparativamente rápido (Brinkman y Leipe, 2001) que, al igual que los métodos de evolución mínima (ver más adelante), descubre la filogenia correcta si las secuencias analizadas cumplen con el supuesto de homogeneidad de tasas de divergencia.

* Métodos de mínimos cuadrados

El método Fitch-Margoliash (FM; Fitch y Margoliash,1967) y otros similares tratan de calcular el valor mínimo posible para la suma de las diferencias cuadradas entre las distancias pareadas observadas y las distancias pareadas esperadas para una topología dada y escogen la topología que minimiza esta diferencia (Nei, 1996). Cavalli-Sforza y Edwards (1967) han sugerido el uso de este método para la obtención de distancias genéticas a partir de datos de frecuencia de genes en poblaciones (Nei, 1996). Un problema de los métodos de mínimos cuadrados es que pueden dar longitudes negativas de ramas lo que hace que la precisión de la topología obtenida no sea muy buena (Nei, 1996;

Figura 14. Descomposición de estrella. Los algoritmos de construcción de árboles tales como *neighbor-joining* o los de mínimos cuadrados, usan este algoritmo una vez que han calculado las distancias entre los terminales. Los terminales más similares se unen y se inserta una rama entre ellos y el resto de la estrella. Subsecuentemente, la nueva rama se consolida de manera que su valor es la media de los dos valores originales de A y B, dando una estrella con *n*-1 terminales. El proceso se repite hasta que sólo queda un terminal (modificado de Brinkman y Leipe, 2001)



Kuhner y Felsenstein,1994; Saitou y Nei, 1987; Sourdis y Krimbas, 1987). Se ha propuesto rectificar esto imponiendo la restricción de que las longitudes de ramas calculadas sean positivas (Cavalli-Sforza y Edwards, 1967; Felsenstein, 1988). Cabe mencionar que el método de *neighbor-joining* también puede producir ocasionalmente árboles con ramas de longitudes negativas.

*Métodos de evolución mínima

Las longitudes de las ramas de un árbol se estiman a partir de la distancia entre todos los pares de secuencias y para todas las posibles topologías se calcula la suma total (*S*) de las longitudes de ramas. El árbol que se escoge al final es aquel que tiene el menor valor de *S*. Las longitudes de ramas pueden estimarse mediante el algoritmo de Fitch y Margoliash (1967) u otro algoritmo de mínimos cuadrados. Rzhetsky y Nei (1993) han mostrado que el árbol que minimiza *S* corresponde a la topología correcta si se utilizan estimadores sin sesgo de las substituciones de aminoácidos o de ácidos nucleicos, lo que implica que este método es válido si hay uniformidad en las tasas de divergencia de

las secuencias (reloj molecular). También han mostrado que el método converge en la filogenia correcta sólo si hay un número suficientemente grande de nucleótidos o aminoácidos.

Otra reserva sobre el método de evolución mínima: tiende a ser oneroso en recursos computacionales, por lo que Rzhestsky y Nei (1994) recomiendan una estrategia en la que primero se construye el árbol de *neighbor-joining* (NJ) y luego un conjunto de topologías cercanas al árbol NJ se evalúan para encontrar una que tenga una *S* aún menor que la del árbol NJ. Este *árbol provisional de evolución mínima* es el punto de inicio para una nueva iteración de este mismo procedimiento que se repite hasta que no se pueda encontrar un árbol con una *S* aún menor y el último árbol encontrado se considera *el* árbol de evolución mínima. Kumar (1996) desarrolló una extensión de este algoritmo que hace una exploración más completa (sin ser exhaustiva) del espacio de árboles posibles incluso cuando el número de secuencias es muy grande.

Métodos basados en caracteres

Los métodos anteriores calculan las distancias entre todos los pares de secuencias representativas de los taxa a estudiar de acuerdo con algún criterio preestablecido y después desechan los datos efectivamente disponibles (Brinkman y Leipe, 2001). En cambio, los métodos basados en caracteres encuentran los árboles que optimizan la distribución de los patrones efectivamente observados *para cada caracter*. De esta forma, las distancias entre cada par de secuencias no están unívocamente determinadas (no son únicas) pues dependen de la topología de la filogenia en la que se midan.

El problema de la reconstrucción filogenética, desde el punto de vista de la máxima parsimonia, consiste en encontrar un árbol bifurcado que tenga las diferentes secuencias observadas asociadas a las puntas de sus ramas (lo que se llama un grafo etiquetado) y que permita postular una historia de ancestría y descendencia entre las secuencias observadas y sus ancestros hipotéticos que requiera tan pocos cambios (substituciones de nucleótidos y aminoácidos) como sea posible (Fitch, 1977), es decir, el árbol más parsimonioso.

Esta forma del problema de la búsqueda del árbol más parsimonioso basada en el uso de datos moleculares difiere de las que usan otro tipo de datos en que no se especifica *a priori* qué estados de caracter son derivados y cuales ancestrales y en que los cambios de estados de caracter no se consideran irreversibles. Es importante notar que a menudo existirá más de un árbol con el

mismo número *mínimo* de cambios de estado de caracter por lo que no será posible inferir un único árbol filogenético (Li, 1997).

El primer paso en la búsqueda de los árboles más parsimoniosos con datos moleculares consiste en definir cuáles son los sitios informativos. Un sitio nucleotídico o aminoácido es informativo si nos permite preferir ciertos árboles sobre otros y esta decisión no podemos tomarla utilizando aquellos sitios en los que sólo hay un tipo de nucleótido (o aminoácido) ni para aquellos sitios en los que hay un solo tipo de nucleótido no singular (un caracter singular es aquel que sólo está representado una vez en un sitio dado, por ejemplo, en la siguiente alineación, la adenina A del sitio 2 es una singularidad y todos los nucleótidos del sitio 4 son singularidades).

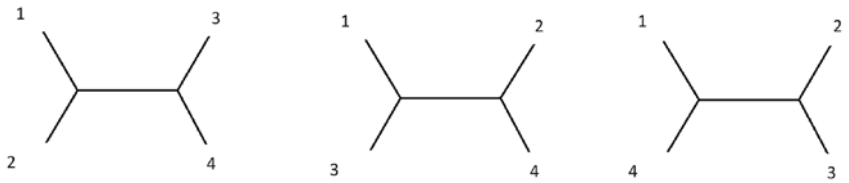
Sitio									
Secuencia	1	2	3	4	5	6	7	8	9
1	A	A	G	A	G	T	G	C	A
2	A	G	C	C	G	T	G	C	G
3	A	G	A	T	A	T	C	C	A
4	A	G	A	G	A	T	C	C	G
					*		*		*

Si no tomamos en cuenta los sitios que sólo contienen un tipo de caracter o un solo tipo que no es singular, nos quedamos sólo con las posiciones que pueden afectar la estuctura del árbol más parsimonioso, es decir, las posiciones informativas o que contienen discordancias. Una discordancia es la presencia en dos taxa de nucleótidos diferentes, no singulares (Fitch, 1977). Otra forma de ver los sitios informativos es pensar que son aquellos en los cuales hay por lo menos dos tipos de nucleótidos y todos los tipos presentes se encuentran en por lo menos dos de las secuencias estudiadas. En la alineación anterior sólo las posiciones marcadas con un asterisco son discordantes e informativas desde el punto de vista de la parsimonia. Cuando estudiamos cuatro unidades taxonómicas operativas, como es el caso en el ejemplo de esta alineación, sólo hay tres árboles posibles que las conecten, los llamaremos I, II y III:

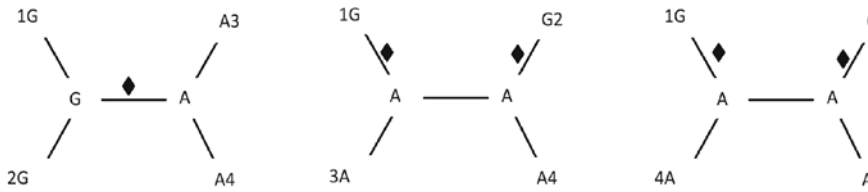
Como hemos visto, estos árboles se pueden representar de una manera más compacta: árbol I ((1,2),(3,4)); árbol II ((1,3),(2,4)); árbol III ((1,4),(2,3)). Obsérvese como se distribuiría el caracter 5 de nuestro ejemplo en cada uno de los tres posibles árboles (postulando estados ancestrales hipotéticos en los nodos internos) (modificado de Li y Graur, 1991):

Para el sitio 5, el árbol I requiere postular una sola substitución (el paso de G a A o bien de A a G en la rama ancestral marcada con un diamante) para explicar la distribución actual de los caracteres, mientras que el árbol II y el árbol III requieren cada uno postular dos sustituciones. El lector debe analizar los caracteres 7 y 9 de esta manera para saber cuál es el árbol que requiere menos cambios para la totalidad de los caracteres informativos.

Cuando se está estudiando más de cuatro unidades taxonómicas operativas la situación es un poco más complicada pero el principio es el mismo: inferir cuál es el mínimo número de sustituciones requeridas por un árbol dado (Li, 1997).



Para poder contar el número de sustituciones a lo largo de un árbol dado se necesita postular las secuencias ancestrales de cada uno de los clados implícitos en ese árbol. Por ejemplo, para el árbol ((((A, B),(C, D)),E),F), G) se requiere postular estados de caracter para el ancestro del clado (A,B), del clado (C, D), del clado ((A, B),(C, D)), del clado (((A, B), (C, D)),E),



etc. Fitch (1971, 1977) propuso un algoritmo que hoy es estándar para inferir dichos estados ancestrales mediante parsimonia. En este algoritmo, un nodo interno representa un ancestro que, a diferencia de las unidades taxonómicas observadas denotadas por las puntas de las ramas, puede tener

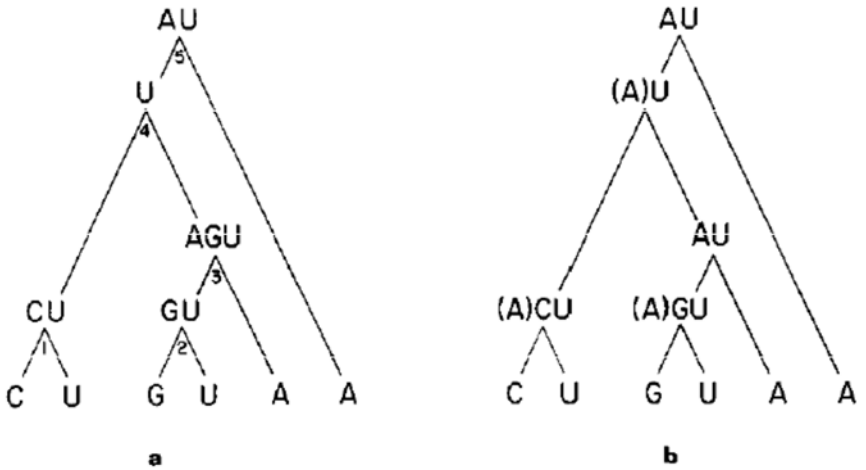
más de un estado de caracter para cada sitio de su secuencia asociada. El conjunto de estados de caracter para un sitio dado de un nodo interno será el resultado de la operación de *intersección* de los conjuntos de sus nodos inmediatamente descendientes *si* dicha intersección no es un conjunto vacío, y el resultado de la operación de *unión* de los conjuntos de sus nodos inmediatamente descendientes si la intersección es vacía. En la figura 15 se representa en las puntas de las ramas el estado de caracter para un solo sitio en una secuencia de ARN en seis taxa.

Al aplicar este algoritmo, cada vez que se necesita usar la operación de unión para formar el conjunto de estados de caracter de un nodo interno, eso implica que evolutivamente debió ocurrir una substitución en esa posición durante el periodo histórico denotado por las ramas que conectan al nodo interno con sus nodos descendientes. Por lo tanto, si contamos el número de uniones obtenemos el número mínimo de substituciones requeridas para explicar la distribución actual de los estados de caracter dado un árbol filogenético. En nuestro ejemplo, este número es cuatro (denotado por los nodos 1, 2, 3 y 5 de la figura 15a). En la figura 15b se muestra la existencia de otros 104 árboles no enraizados y 944 enraizados para conectar seis OTUs, aquí hay dos de ellos (modificado de Li, 1997).

Cada uno de estos árboles requiere tres uniones para formar los conjuntos de los nodos internos e implican tres substituciones. Precisamente, ése es el número mínimo de substituciones requeridas para explicar las diferencias entre los nucleótidos en la posición que estamos estudiando porque en ella se presentan los cuatro tipos de nucleótidos, por lo que a partir de un solo nucleótido ancestral deben haber ocurrido por lo menos tres cambios mutacionales. Existen muchos otros árboles que requieren tres substituciones, por lo tanto, un sitio informativo puede favorecer a muchos árboles alternativos, a diferencia de lo que ocurre cuando tenemos sólo cuatro OTUs (Li, 1997). Aunque es fácil inferir el número de substituciones mínimas requeridas, estimar el orden exacto en que ocurrieron los cambios es a menudo más difícil (véase Fitch 1971, 1977 para el procedimiento).

El procedimiento de estimar el número de sustituciones requeridas por un árbol dado debe aplicarse a todos los árboles que se estén considerando. En la práctica, para más de 20 OTUs, no es factible evaluar todos los árboles posibles (exploración exhaustiva) por lo que es necesario recurrir a exploraciones heurísticas en las que se evalúa un subconjunto de todos los árboles posibles con la esperanza de que en ese subconjunto se encuentre el o los árboles óptimos (Page y Holmes, 1998).

Figura 15. (a) uno de los 945 posibles árboles enraizados que conectan seis OTUs (nótese que el algoritmo de Fitch no exige árboles enraizados); en (b) se muestra que se puede incluir a A como posible nucleótido ancestral en los nodos 1, 2 y 4 además de los que se indican en (a) porque A es un posible nucleótido ancestral común (nodo 5) de las seis secuencias (Li, 1997; figura de Fitch, 1971)



Máxima verosimilitud

El concepto de máxima verosimilitud se atribuye generalmente a Ronald Fisher quien lo describió en 1922 (Fisher 1922, 1950). Depende de una completa especificación de los datos y de un modelo probabilístico que describa los datos. La probabilidad de observar los datos dado el modelo que se ha tomado cambiará dependiendo de los valores que tomen los parámetros del modelo. En la estimación por máxima verosimilitud se trata de escoger el valor de un parámetro que maximiza la probabilidad de observar los datos.

Para familiarizarnos con el concepto de verosimilitud pensemos en el paradigma por excelencia de la probabilidad: echar volados. Con una moneda buena la probabilidad de obtener águila o sol es, en ambos casos, de 0.5, pero supongamos que queremos averiguar si la moneda es buena por lo que hay que estimar la probabilidad de cada resultado. En la aproximación de máxima verosimilitud los elementos son (a) los datos, (b) un modelo que describa la probabilidad de observar los datos que tenemos y (c) el criterio de máxima verosimilitud (Huelsenbeck y Crandall, 1997). El modelo que describe la pro-

babilidad de observar *h* águilas en *n* lanzamientos es la distribución binomial que tiene la siguiente forma:

$$\Pr [h \mid p,n] = \binom{n}{h} p^h (1-p)^{n-h}$$

donde *p* es la probabilidad de obtener águila, el coeficiente binomial $\binom{n}{h}$ da el número de *h* resultados exitosos en *n* pruebas (de lanzamientos) y la barra vertical significa "dando por supuesto".

$\Pr [h \mid p,n]$ debe de leerse "probabilidad de *h* dados *p* y *n*".

Si los resultados individuales son independientes entre sí, la función de verosimilitud es simplemente la probabilidad conjunta de observar los datos bajo el modelo supuesto. En el caso de los volados, bajo la distribución binomial, esta función es:

$$L(p \mid h, n) = \binom{n}{h} p^h (1-p)^{n-h}$$

A menudo se prefiere usar el logaritmo de la verosimilitud para facilitar los cálculos sin que esto afecte cuál es el valor de *p* que maximiza la verosimilitud.

Supongamos que en 10 lanzamientos obtenemos seis águilas y cuatro soles. Si graficamos *L* como función de *p* en este caso observaremos que la verosimilitud se maximiza en donde *p* corresponde a la proporción de veces que obtuvimos águila en nuestro experimento, es decir en 0.6. De hecho esta es una forma numérica de obtener un estimado de máxima verosimilitud del valor de *p*. Este valor también se puede encontrar analíticamente buscando el punto en el que la pendiente de la derivada de la función de verosimilitud respecto a *p* es cero. Si lo hacemos encontraremos que este estimado es $p = {}^{h}/_{n}$ o sea que en este caso el estimado de *p* es simplemente la proporción de águilas observadas en nuestro experimento.

Matemáticamente, los estimados de máxima verosimilitud se consideran consistentes y eficientes (es decir, convergen al valor correcto a medida que crece el número de datos y tienen poca varianza) pero a menudo no son libres de sesgo (por ejemplo, el estimado de máxima verosimiltud de la varianza de la población en una distribución normal es sesgado; Huelsenbeck y Crandall, 1997).

Vale la pena subrayar la distinción entre probabilidad y verosimilitud. Pr(O|H) es la probabilidad de las observaciones O dada la hipótesis H; si

se considera como una función de O, define una distribución estadística discreta o continua. Así, si sumamos (o integramos) sobre todas las posibles observaciones O, obtenemos la unidad. Por otra parte, la verosimilitud está basada en datos fijos O y se puede ver como una función de las hipótesis o de los parámetros cuando estos y aquellas varían. Pero esta función de ninguna manera origina una distribución estadística y nada en su definición implica que si se sumara sobre todas las hipótesis posibles (si tal cosa fuera posible) o se integrara sobre todos los posibles valores de parámetros, se obtendría algún valor en especial (particularmente, nada indica que se obtendría la unidad). No hay ningún significado especial asociado a ninguna parte del área bajo la curva de verosimilitud, o a la suma de las verosimilitudes de dos o más hipótesis (Edwards, 1992).

Conceptualmente, la aplicación de la máxima verosimilitud en filogenética es tan sencilla como en el caso del volado. El siguiente ejemplo está tomado de Huelsenbeck y Crandall (1997). Comencemos con un alineamiento de secuencias de ADN de s=4 taxa:

Taxon 1 ACCAGC
Taxon 2 AACAGC
Taxon 3 AACATT
Taxon 4 AACATC

las observaciones son los patrones en cada sitio (columna) del alineamiento, por ejemplo, $\mathbf{x}_1$= {A,A,A,A} y $\mathbf{x}_6$= {C,C,T,C} para la primera y la sexta columnas. La muestra consiste en n vectores (tantos como haya sitios en el alineamiento) similares a los vectores $\mathbf{x}_1$ y $\mathbf{x}_6$. En el alineamiento hay dos vectores ($\mathbf{x}_1$ y $\mathbf{x}_4$) que contienen el mismo patrón. Cuando estudiamos alineamientos de nucleótidos, existe un total de $r=4^s=256$ patrones posibles para s especies (o unidades taxonómicas operacionales, en general). Como hay un número finito de patrones posibles podríamos representar nuestros datos como el número de veces que observamos cada uno de estos patrones, entonces nuestra matriz de datos original se convertiría en:

Taxon 1	A A A A A A A A A...C...C...C...G...TTT
Taxon 2	A A A A A A A A A...A...C...C...G...TTT
Taxon 3	A A A A C C C C G...A...C...T...T... TTT
Taxon 4	A C G T A C G T A...A...C...C...T...CGT
Número	2 0 0 0 0 0 0 0 0 0... 1... 1...1...1... 000

y ahora tenemos una matriz de 4 por 256 con los 256 patrones posibles numerados 1,2,…,*r*. La mayor parte de los patrones posibles no se observan en nuestros datos, lo cual se refleja en que la línea que indica el número de veces que se observó un patrón contiene muchos ceros, pero sí se observan cinco de los patrones (que ahora llamamos y_1= {A,A,A,A}, y_{65}= {C,A,A,A}, y_{86}= {C,C,C,C}, y_{94}= {C,C,T,C} y y_{176}= {G,G,T,T}), la cantidad de sitios en que se presenta cada uno de los patrones está contenida en el vector n (n_1=2, n_{65}=1, n_{86}=1, n_{94}=1, n_{176}=1, y todos los demás n_i=0).

La máxima verosimilitud presupone un modelo explícito para los datos (por ejemplo, en el caso de los volados, se presuponía una moneda sin sesgo, con la misma probabilidad de dar águila que sol). En este ejemplo, como en el de la moneda, se considera que los datos son variables aleatorias, pero en vez de dos resultados posibles, ahora hay $r = 4^s$ posibles resultados en el caso de secuencias de ADN. Por lo tanto, los datos se pueden describir con una distribución multinomial (una generalización de la binomial) que tiene la siguiente forma:

$$\Pr[n_1, n_2, \ldots, n_r \mid p_1, p_2, \ldots p_r] = \binom{n}{n_1, n_2, \ldots, n_r} \prod_{i=1}^{r} p_i^{n_i}$$

en donde el primer factor del lado derecho es el número de maneras en que *n* objetos se pueden agrupar en *r* clases, n_i es el número de observaciones del *i*-ésimo patrón y p_i es la probabilidad de que ocurra el patrón *i*. Un estimado de máxima verosimilitud de p_i es ${}_{est}p_i = n_i/n$ (es decir, la probabilidad de la clase *i*-ésima es la proporción del tiempo que fue observada). La verosimilitud puede entonces calcularse, suponiendo una distribución multinomial, igualando la verosimilitud a la ecuación anterior. Sin embargo, al usar la distribución multinomial no se puede estimar la topología ni otros parámetros biológicos, de manera que para poder estudiarlos se presuponen modelos que incorporan información filogenética.

Al igual que el modelo de probabilidad multinomial, los modelos filogenéticos especifican al probabilidad de observar diferentes patrones. Un modelo filogenético de datos moleculares debe de incorporar al menos un árbol (τ) que conecte a las secuencias así como longitudes (v) de ramas, expresadas en términos de número esperado de cambios por sitio y un modelo de cambio para las secuencias. Considérese sólo uno de los patrones de sitio de nucleótidos (el y_{176}) para el árbol de la figura 16 (véase página siguiente).

Como no conocemos la identidad de los nucleótidos en los nodos internos *i*, *j* y *k*, la probabilidad de observar el patrón de sitio 176 es la suma de 64 términos (las $4^3 = 64$ posibles asignaciones de nucleótidos a los nodos *i, j* y

k). Esta probabilidad se puede expresar:

$$\Pr [\, y_{176} = \{G, G, T, T\} \mid \tau, v1, \ldots v_6, \Theta]$$

$$= \sum_{i=1}^{4} \sum_{j=1}^{4} \sum_{k=1}^{4} \pi_\kappa \, p_{Gi}(v_1, \Theta) \, p_{ik}(v_5, \Theta) \, p_{jk}(v_6, \Theta)$$

donde se asume que los nodos internos tienen los nucleótidos A, C, G o T si los nodos *i, j* o *k* son iguales a 1, 2, 3 o 4 respectivamente, π_i es la frecuencia en el equilibrio del nucleótido *i* y $p_{ij}(v_i, \Theta)$ es la probabilidad de observar los nucleótidos *x* e *y* en los dos extremos de una rama dada la longitud de la rama y otros parámetros, Θ, del modelo de sustitución. Si en lugar de nucleótidos se están considerando aminoácidos la ecuación involucra la suma sobre $20^3 = 8000$ posibles asignaciones de aminoácidos en los nodos internos.

Si suponemos que cada sitio evoluciona independientemente de los otros, la verosimilitud de un árbol τ es:

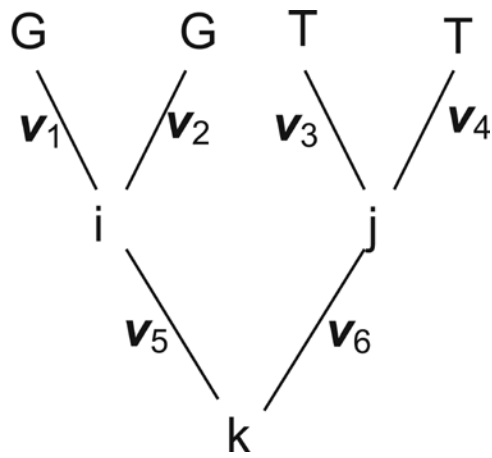
$$L(\tau, v, \Theta \mid y_1, \ldots, y_r) = \binom{n}{n_1, n_2, \ldots, n_r} \prod_{i=1}^{r} \Pr[y_i \mid \tau, v, \Theta]^{n_i}$$

donde v es un vector que contiene las longitudes de las ramas y *r* es el número de patrones posibles para *s* secuencias. El método de Felsenstein (1981) consiste en escoger el árbol que maximice la verosimilitud como el mejor estimador de la filogenia. En principio, para encontrar dicho árbol se debe hacer el cálculo para cada uno de los (2*s*-5)! árboles no enraizados posibles, pero afortunadamente existen métodos que calculan la verosimilitud sumando todas las posibles asignaciones en los nodos internos (Felsenstein 1981), optimizando las longitudes de las ramas (Lewis, *et al.*, 1996) o bien explorando el universo de los árboles (Swofford *et al.*, 1996). Además, recientemente se han desarrollado programas de cómputo que pueden calcular eficientemente árboles de máxima verosimilitud para conjuntos de datos de cientos de secuencias y miles de sitios (Guindon y Gascuel, 2003; Stamatakis *et al.*, 2004).

Para calcular la probabilidad de observar un determinado patrón de sitios hay que especificar las probabilidades de transición $p_{ij}(v_i, \Theta)$. Imaginemos un caso muy simple en el que sólo existen dos estados de caracter (0 y 1). La tasa de cambio de 0 a 1 o de 1 a 0 en una cantidad infinitesimal de tiempo δ*t* está descrita por la matriz de tasas Q

$$Q = \{q_{ij}\} = \begin{pmatrix} -\lambda\pi_1 & \lambda\pi_1 \\ -\lambda\pi_0 & \lambda\pi_0 \end{pmatrix}$$

Figura 16. El método de máxima verosimilitud supone que las secuencias observadas (aquí, el patrón 176 del ejemplo del texto) se relacionan mediante un árbol (τ) con longitudes de ramas (v_1,…., v_6) especificadas en términos de sustituciones esperadas por sitio. La probabilidad de observar los datos (aquí, el patrón y_{176}) es la suma sobre todas las posibles asignaciones de nucleótidos a los nodos internos *i, j* y *k*. (Huelsenbeck y Crandall, 1997)



donde λ es la tasa de cambio de 0 a 1 o de 1 a 0 y π_0 y π_1 son las frecuencias en el equilibrio de 0 y 1 respectivamente. Para calcular la probabilidad de observar un cambio en un intervalo arbitrario de tiempo se hace el siguiente cálculo: $\mathbf{P}(t, \Theta) = p_{ij}(t, \Theta) = e^{Qt}$ (Cox y Miller, 1977). El vector Θ contiene los parámetros del modelo de substitución (en este caso $\Theta = \{\pi_0 , \pi_1 \}$). Para muchos modelos de substitución la matriz de probabilidades de transición $\mathbf{P}(t, \Theta)$ se puede calcular analíticamente, por ejemplo, las probabilidades para el caso de dos estados son:

$$P(t, \Theta) = \{p_{ij}(t, \Theta)\} = \begin{pmatrix} \pi_0+(1-\pi_0)e^{-\lambda t} & \pi_1-1-\pi_1 e^{-\lambda t} \\ \pi_0 - \pi_0 e^{-\lambda t} & \pi_1+(1-\pi_1)e^{-\lambda t} \end{pmatrix}$$

Sin embargo, para matrices de tasas más complicadas, la matriz de probabilidades se tiene que calcular con métodos numéricos (Huelsenbeck y Crandall, 1997).

Como ya se mencionó, para encontrar el árbol que tenga la máxima verosimilitud, se debería, en principio, explorar la totalidad del universo de los

árboles asociados a determinado número de taxa, pero al igual que sucede con otros métodos esto es prácticamente imposible, incluso para un número moderado de taxa (Swofford *et al*., 1996). Como la verosimilitud de un árbol no depende sólo de la topología sino también de valores de parámetros que no siempre se pueden encontrar analíticamente, incluyendo la longitud de las ramas, los problemas de cómputo en la reconstrucción por máxima verosimilitud son particularmente difíciles (Guindon y Gascuel, 2003), por lo que los métodos prácticos se basan en aproximaciones heurísticas. Por ejemplo, programas populares como PAUP* (Swofford, 2000) o PHYLIP (Felsenstein, 1999) buscan los árboles con las más grandes verosimilitudes mediante el algoritmo de 'trepar colinas' en el que primero se van añadiendo uno a uno los taxa al árbol de acuerdo con determinada regla (por ejemplo, la adición de ramas, *stepwise addition*, combinada con el intercambio de ramas, *branch swapping*; Nei y Kumar, 2000) y evaluando el efecto sobre la verosimilitud al mismo tiempo que se optimizan las longitudes de las ramas. Cuando la posición de inserción de una rama es la mejor entre todas las posibles (en la etapa de agregación de ramas) o cuando un rearreglo produce un árbol más verosímil (en la etapa de rearreglos) el árbol producido se toma como punto de partida para la siguiente iteración. Existen programas que implementan nuevas versiones del algoritmo de trepar colinas y logran velocidades de cálculo similares a las de métodos basados en distancia o parsimonia (Stamatakis *et al*., 2004; Guindon y Gascuel, 2003; Lemon y Milinkovitch, 2002; Ranwez y Gascuel, 2002; Adachi y Hasegawa, 1996; Olsen *et al*., 1994). Por ejemplo, el programa PHYML (Guindon y Gascuel, 2003) ajusta simultaneamente y progresivamente la topología y las longitudes de ramas de manera que la verosimilitud se incrementa en cada iteración hasta que se alcanza un óptimo.

Aproximaciones bayesianas

Las aproximaciones bayesianas se basan en una cantidad llamada la probabilidad posterior de un árbol, una medida que puede ser interpretada como la probabilidad de que un árbol dado sea el que describe correctamente la historia de la muestra taxonómica que estamos estudiando (Huelsenbeck *et al*., 2001). La probabilidad posterior es proporcional al producto de la probabilidad *a priori* y de la verosimilitud (Holder y Lewis, 2003). Más concretamente, la probabilidad posterior está dada por el resultado algebraico a veces pomposamente llamado *teorema de Bayes* que podemos expresar en términos de filogenias como que la probabilidad posterior de un árbol *i*,

incluyendo las longitudes de sus ramas y un modelo de substituciones, dados los datos, es decir el alineamiento (todo lo cual podemos notar como Pr[árbol *i* | datos]), es igual al producto de la probabilidad de los datos dados el árbol, las longitudes de ramas y el modelo de substitución multiplicada por la probabilidad *a priori* del árbol y el todo dividido por la probabilidad de los datos. La probabilidad *a priori,* o previa, de un árbol *i* (que podemos notar Pr[árbol *i*]) es la probabilidad que le asignamos antes de hacer cualquier observación. Como casi siempre se considera que los árboles son *a priori* equiprobables, para una alineación de tres secuencias, cada uno de los tres posibles árboles que las conecten tendrá una probabilidad previa de 0.33. La probabilidad de los datos dados el árbol *i*, las longitudes y el modelo de substitución (Pr[datos | árbol *i*]) requiere hacer suposiciones específicas sobre el proceso que generó los datos. Después de haber leído la sección anterior es fácil darse cuenta que la probabilidad de los datos dados el árbol y los parámetros es, de hecho, la verosimilitud de la alineación.

Los métodos bayesianos de reconstrucción filogenética están cercanamente emparentados a los de máxima verosimilitud; son similares a estos en que también requieren plantear un modelo explícito de cambio de caracteres y en que los 'mejores' árboles son aquellos que son consistentes tanto con ese modelo como con los datos del alineamiento. Por otra parte existe una diferencia sutil entre los métodos de máxima verosimilitud y los bayesianos pues mientras aquellos buscan el árbol que maximiza la probabilidad de observar los datos –es decir, el alineamiento—dado el modelo de substitución y dicho árbol, los métodos bayesianos consideran que el mejor estimado de la filogenia es aquel árbol que tiene la máxima probabilidad de ser observado dados un modelo de substitución y el alineamiento. Otra diferencia importante es que la máxima verosimilitud no trata a las variables del modelo como variables aleatorias mientras que en el análisis bayesiano todo, tanto datos como parámetros, es considerado como variables aleatorias, lo que significa que la incertidumbre acerca de tal o cual parámetro está descrita por la distribución de probabilidad posterior.

Aunque la probabilidad posterior es fácil de formular, calcular la probabilidad posterior de una filogenia es computacionalmente oneroso pues requiere hacer sumas de todos los árboles posibles y, para cada árbol, integrar sobre todas las posibles combinaciones de longitudes de ramas y todos los posibles valores de los parámetros de los modelos de substitución (Huelsenbeck *et al*., 2001). Afortunadamente existen varios métodos numéricos para obtener una aproximación de la probabilidad posterior de un árbol. El objetivo de estos métodos es análogo a lo que se intenta hacer en los métodos de máxima

verosimilitud, al usar algoritmos como el de trepar colinas, es decir, se trata de encontrar, mediante un proceso que involucra ensayo y error, los árboles óptimos o cercanos a los óptimos. Probablemente el método más útil de optimización numérica en inferencia filogenética bayesiana sea el método de Monte Carlo con cadenas de Markov (MCMC por sus siglas en inglés; Huelsenbeck *et al*., 2001). La idea del algoritmo MCMC es que en cada iteración del algoritmo se "propone" un nuevo árbol (lo que incluye variaciones en la topología, los parámetros del modelo de sustitición y el vector de longitudes de ramas). El nuevo árbol se obtiene por perturbación de un árbol que se había propuesto previamente, por lo que es probable que ambos sean más o menos parecidos. Si el nuevo árbol tiene una mejor probabilidad posterior que el viejo, se toma como punto de partida para la siguiente iteración (es "aceptado"); si, en cambio, el nuevo árbol tiene una menor probabilidad posterior que el viejo, sólo es aceptado si el cociente entre su probabilidad y la del árbol actual es mayor que un número tomado al azar en el intervalo 0,1 (Lewis, 2001b; Lewis y Holder, 2003). Si los primeros desplazamientos en el espacio de los árboles se descartan, la distribución de los puntos visitados es una buena aproximación de las regiones con árboles de mayor probabilidad posterior. De hecho se dice, no sin ironía, que este algoritmo garantiza encontrar el árbol de mayor probabilidad posterior si se ejecuta durante un tiempo infinito (P. Vinuesa, com. pers.). Esto apunta hacia una dificultad potencial del algoritmo consistente en que sólo da buenos resultados si se deja correr lo suficiente, pero es difícil saber de antemano cuanto tiempo es suficiente. Afortunadamente, las implementaciones más recientes del MCMC bayesiano en reconstrucción filogenética (notablemente el programa MrBayes 3.01 [Ronquist y Huelsenbeck, 2003]) La ventaja es que si guarda un registro de los árboles visitados se puede saber en que regiones del espacio de los árboles ha pasado más tiempo el algoritmo de exploración, lo cual es un indicador de cuáles son los árboles (y también los subclados) que están mejor soportados estadísticamente, y esta información se obtiene al mismo tiempo que se está buscando los árboles óptimos.

Otra ventaja de los métodos bayesianos es que permiten usar modelos evolutivos muy complejos, que incluyan, por ejemplo, combinaciones de secuencias biológicas cada una de las cuales puede estar sujeta a circunstancias evolutivas diferentes e incluso permiten analizar conjuntamente datos de secuencias moleculares (de ácidos nucléicos y aminoácidos), datos de patrones obtenidos con enzimas de restricción y datos morfológicos (Lewis, 2001b; Holder y Lewis, 2003).

Los métodos bayesianos comparten con los métodos de máxima verosimilitud otro aspecto atractivo: su robustez frente al artefacto conocido como "atracción de ramas largas". Imaginemos a las unidades taxonómicas A, B, C y D representadas por secuencias moleculares; las secuencias de B y D son relativamente parecidas entre ellas pero las secuencias de A y C son, cada una, bastante diferentes de las otras tres secuencias. En un caso así se dice que en cualquier árbol que conectara a las cuatro secuencias las ramas que conectaran tanto a A como a C serían ramas largas. Ahora bien, como vimos anteriormente, cuando las secuencias están conectadas al árbol por ramas largas es más probable que tengan similitud homoplásica con otras secuencias que también están conectadas por ramas largas y en esas circunstancias, un método que presuponga que todos los tipos de substituciones de caracteres son igualmente probables (como, por ejemplo, la máxima parsimonia) tenderá a agrupar a los taxa A y C por un lado y a los taxa B y D por otro (incluso si ((A,B),(C, D)) es el árbol correcto) y, en general el efecto será que las secuencias de ramas largas parecerán atraerse entre sí. En contraste con esto, los métodos que plantean un modelo explícito de substituciones de cracteres en el que diferentes tipos de substitución tengan diferentes probabilidades (como los métodos de máxima verosimilitud y los bayesianos) son mucho más robustos frente a la atracción de ramas largas porque sus cálculos incorporan las probabilidades de todos los cambios posibles, incluyendo la posibilidad de que las similitudes entre secuencias sean homoplásicas, y ponderan todas las posibilidades de acuerdo al modelo de substituciones bajo el cual se plantea que están evolucionando las secuencias. Sin embargo, hay que enfatizar que esta robustez sólo existe si el modelo de substitución supuesto es el correcto o por lo menos tiene altas probabilidades de ser cercano al correcto, de otra forma los métodos de máxima verosimilitud y bayesianos corren el riesgo de ser incluso más engañosos que los de máxima parsimonia. Aquí la regla pragmática es que si no se sabe cuáles son los modelos de substitución de mejor describen a las secuencias con las que trabajamos es mejor hacer una aproximación con máxima parsimonia, y que si se plantea usar los métodos que suponen modelos probabilísticos hay que hacer todo esfuerzo por encontrar el modelo que con mayor probabilidad describe a nuestros datos.

Evaluación estadística de árboles

Bootstrap no paramétrico

Una vez que se obtiene una estimación de la filogenia que mejor describe las relaciones entre nuestros datos, la pregunta que surge inmediatamente es ¿qué tan bien soportan los datos las relaciones que se muestran en el árbol? (Holder y Lewis, 2003). Una de las formas más populares de evaluar la confiabilidad de una inferencia filogenética es el análisis de *bootstrap* (Efron, 1979; Felsenstein, 1985). El primer paso en un análisis de *bootstrap* es remuestrear, con reemplazo, las columnas del alineamiento de secuencias, es decir, en el alineamiento remuestreado ("de *bootstrap*") determinada columna del alineamiento original puede aparecer dos o más veces, mientras que otras columnas pueden no estar representadas para nada.

Si se reconstruye un nuevo árbol con el alineamiento remuestreado y se obtiene un árbol similar al árbol obtenido con los datos originales, eso incrementa nuestra confianza en la reconstrucción inicial; si, en cambio el nuevo árbol se ve muy diferente del árbol original eso puede ser una indicación de que este es poco confiable (o quizá más exactamente, de que la información evolutiva está distribuida en forma muy heterogénea en los datos del alineamiento). Al remuestrar numerosas veces (digamos, unas mil veces, generando mil alineamientos de *bootstrap*) es posible asignar intervalos de confianza a cada rama interna del árbol. Si por ejemplo, ciertas secuencias aparecieron formando un grupo monofilético en el 74% de los árboles generados a partir de los remuestreos, diríamos que ese agrupamiento tiene un soporte por *bootstrap* de 74% y nuestra confianza en la monofilia de ese grupo será mayor que la de un grupo con un soporte de, digamos, 25%. En general, el *bootstrap* se considera una medida conservadora de la confiabilidad (Hillis y Bull, 1993; Zarkhikh y Li, 1992; Holder y Lewis, 2003; ver también Felsenstein y Kishino, 1993; Efron *et al*., 1996) por lo que un valor de *bootstrap* de 70% o más puede ser indicador de buen soporte y las afirmaciones basadas en agrupamientos con menos de 50% de soporte deberían hacerse con cautela.

La probabilidad posterior bayesiana

La inferencia bayesiana toma una postura diferente frente al problema de la confiabilidad de las reconstrucciones filogenéticas. En lugar de buscar un único árbol que represente la filogenia óptima o correcta, se busca muestrear toda la

distribución de probabilidad posterior de las filogenias posibles y para lograrlo los árboles se muestrean de forma proporcional a su probabilidad posterior (recuérdese que el algoritmo MCMC pasa más tiempo en la regiones del espacio de parámetros correspondientes a los árboles con mayor verosimilitud o mayor probabilidad posterior). Una vez que se tiene una muestra con esas características, es posible discernir los elementos comunes a los árboles con mayor probabilidad posterior. Por ejemplo, se puede construir un árbol de consenso de los árboles de la muestra y en él se puede indicar el número de veces que apareció determinado agrupamiento en los árboles de la muestra (es decir, la probabilidad posterior de ese agrupamiento). Esto es *grosso modo* equivalente a hacer un análisis de *bootstrap* en un análisis de parsimonia o máxima verosimilitud, pero es mucho más rápido (Huelsenbeck *et al*., 2001; Larget y Simon, 1999; Yang y Rannala, 1997). Algunos autores han indicado que la credibilidad de los clados medida mediante la probabilidad posterior es demasiado laxa (Suzuki *et al*., 2002) pero es probable que ambos estadísticos estén midiendo cosas diferentes y no sean directamente comparables (Douady *et al*., 2003.)

Existen otras pruebas para evaluar la confiabilidad de las reconstrucciones filogenéticas que son un poco menos conocidas que el bootstrap y la probabilidad posterior y evalúan aspectos ligeramente diferentes. Algunos de estos métodos se han aplicado principalmente en el contexto de la máxima parsimonia; por ejemplo la prueba de permutación (Archie, 1989; Siebert, 1992; Morrone, 2000) evalúa si los datos tienen más estructura jerárquica de la que se esperaría por azar, mientras que la medida de soporte de ramas propuesta por Bremer (1988, 1994; Morrone 2000) trata de determinar que tan bien soportado está un grupo explorando si ese grupo también ocurre en árboles ligeramente menos parsimoniosos. La diferencia en longitud entre los árboles más cortos que incluyen al grupo y los árboles más cortos que no lo incluyen (el número de pasos extra que se necesitan para colapsar al grupo) es lo que se llama índice de decaimiento o soporte de Bremer. El soporte total de un árbol es la suma de todos los índices de decaimiento de ese árbol.

Bootstrap paramétrico

Cuando hemos tratado de determinar cuál es el modelo de substitución que mejor se ajusta a nuestros datos es posible aplicar una versión del *bootstrap* que aprovecha esa información: el *bootstrap* paramétrico. En esta prueba, en lugar de generar pseudorréplicas de nuestros datos mediante

remuestreo con substitución, generamos réplicas simuladas basadas en la probabilidad de que ocurra cada una de las substituciones posibles de acuerdo a nuestro modelo. Estos *sets* de datos simulados se analizan después de forma similar a las pseudorréplicas del *bootstrap* no paramétrico (Huelsenbeck *et al*., 1996)

Árboles aleatorizados (skewness test)

Varios autores han sugerido que la forma de la distribución de las longitudes de los árboles que se pueden generar con un alineamiento dado es un buen indicador de la presencia de señal filogenética de buena calidad en dicho juego de datos (ver Hillis y Huelsenbeck, 1992). La idea básica es que si el árbol más corto (en el sentido de más parsimonioso) no se encuentra significativamente a la izquierda de la distribución de las longitudes de los árboles posibles, existen pocas razones para esperar que un análisis filogenético revelará información sobre las relaciones históricas de los taxa estudiados. Hillis y Huelsenbeck (1992) han preparado un estadístico para medir el grado de sesgo en la distribución de las longitudes de los árboles y tablas de valores críticos para determinar la presencia de señal filogenética de buena calidad en un alineamiento dado.

Pruebas para determinar si hay diferencias significativas entre filogenias alternativas

Kishino y Hasegawa (1989) y Hasegawa y Kishino (1989) (ver Goldman *et al*., 2000) introdujeron métodos que sirven para determinar si las verosimilitudes de dos topologías alternativas para un juego de datos son significativamente diferentes, lo que nos puede ayudar a decidir, entre otras cosas, si la topología de mayor verosimilitud es realmente una estimación mucho mejor de la historia de los taxa estudiados que otras topologías. Goldman y colaboradores (2000) introdujeron mejoras a los métodos de Hasegawa y Kishino e indican cuáles son los potenciales errores a evitar a l usar estas pruebas. Este tipo de pruebas está disponibles en *software* comercial o disponible gratuitamente (P. ej. PAUP* [Swofford, [2000], PAML [Yang, 1997] o HYPHY [Kosakovsky-Pond *et al*., 2005]).

Métodos que pueden colocar OTUs contemporáneos en nodos internos y otras novedades

Los métodos que hemos mencionado se basan en supuestos acerca de los procesos evolutivos que permiten simplificar el análisis de los datos y la interpretación de los resultados, pero que también representan, en cierto sentido, simples caricaturas de la realidad que, seguramente, es mucho más compleja que los supuestos. Por ejemplo, hemos visto que los modelos de substitución de nucleótidos se han vuelto cada vez más complejos en un intento de capturar tanta información como sea posible acerca de la evolución de las secuencias pero incluso los modelos más complejos desarrollados contienen sólo una fracción de los elementos que pueden afectar dicha evolución. El buen uso de los métodos de reconstrucción filogenética y de los modelos evolutivos asociados a ellos depende de estar conscientes de cuáles son sus supuestos, a veces sólo implícitos, que es lo mismo que decir cuáles son sus alcances y cuáles sus limitaciones.

Entre los supuestos implícitos de los métodos de reconstrucción filogenética que hemos visto están la idea de que el proceso evolutivo es esencialmente bifurcante y la de que no podemos saber cómo eran los ancestros de los taxa cuya filogenia queremos reconstruir. La primera idea implica sobre todo una visión en la cual las unidades taxonómicas se forman por la partición de unidades taxonómicas ancestrales (por ejemplo, dos poblaciones actuales son consecuencia de la separación de una población ancestral y las especies actuales se derivan de la separación y posterior diferenciación de grupos que originalmente pertenecían a la misma especie). En esta visión una vez que los grupos se separan no se vuelven a unir, es decir ningún taxón actual sería el producto de la fusión de otros taxa. El árbol filogenético, con sus ramas siempre bifurcadas, es el icono apropiado de este supuesto metodológico. No obstante, sabemos que las unidades taxonómicas a veces son consecuencia de la fusión de otras unidades, existen especies viables que son el producto de la hibridación y las poblaciones dentro de una misma especie guardan el potencial de fusionarse entre sí. Incluso unidades taxonómicas muy remotamente emparentadas podrían intercambiar partes sustanciales de su información genética. Además, incluso si no hay hibridación o transferencia genética horizontal, el hecho de representar la información filogenética como un único árbol tiende a hacernos perder de vista el conflicto que puede existir entre diferentes caracteres de un mismo alineamiento, algunos de

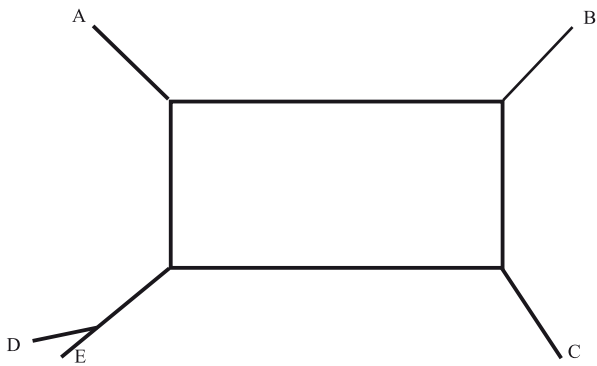
los cuales podrían constituir evidencia a favor de ciertos árboles mientras que otros caracteres podrían apoyar a otros árboles, pero esta información no queda representada en un solo árbol. Los métodos que se mencionan a continuación intentan resolver estas limitaciones.

Descomposición por biparticiones

Este método se basa en el concepto de *splits* o biparticiones, que consisten en separar a los taxa estudiados en dos grupos mutuamente excluyentes. Las biparticiones informativas son aquellas que indican agrupamientos filogenéticos no triviales, es decir aquellas en las que ambos grupos contengan al menos dos taxa. Por ejemplo, si tenemos los taxa A, B, C, D y E, las biparticiones triviales son {{A},{B,C,D,E}}, {{B},{A,C,D,E}}, {{C},{A,B,D,E}}, {{D},{A,B,C,E}} y {{E},{A,B,C,D}}, pero entre las biparticiones no triviales están, {{A,B,C},{D,E}}, {{A,C},{B,D,E}} y {{A,B},{C,D,E}}. Nótese que las dos últimas biparticiones son incompatibles entre sí, es decir, no puede haber un árbol en el que estén agrupados los taxa A y B a la exclusión de los demás taxa y además estén agrupados los taxa A y C a la exclusión de los demás taxa (Page y Holmes, 1998). Ahora bien, es posible que algunos de los caracteres de un alineamiento le den más credibilidad a un árbol que incluya la bipartición {{A,C},{B,D,E}} mientras que un número igual o casi igual de caracteres soporten a los árboles que contengan la bipartición {{A,B},{C,D,E}}, entonces, si elegimos que nuestros datos queden resumidos por los árboles que contienen aquella bipartición, por ejemplo el árbol ((A,C),(B,(D,E))), perderemos la información de que los datos también favorecían en cierto grado los árboles con la bipartición {{A,B},{C,D,E}}, por ejemplo ((A,B),(C,(D,E))). Sin embargo, podemos recurrir a otro tipo de representación para mostrar la información de que los datos favorecen a biparticiones incompatibles entre sí. La figura 17 muestra un diagrama de descomposición en biparticiones en el que se muestra la incompatibilidad entre las dos biparticiones de nuestro ejemplo.

Aunque los diagramas de descomposición de biparticiones (también llamados "redes") pueden ser difíciles de leer al principio, logran representar más información contenida en los datos que los árboles filogenéticos y ofrecen interesantes posibilidades para el análisis exploratorio de datos moleculares especialmente en escalas temporales cortas (Page y Holmes, 1998. Para una introducción a la descomposición de biparticiones ver Bandelt y Dress, [1992] y Dopazo *et al*., [1993]).

Figura 17. Diagrama de descomposición en biparticiones



Una de las razones por las que puede existir conflicto entre diferentes caracteres es la transferencia horizontal de genes, un fenómeno que puede ser extremadamente común en la ecología microbiana y sin duda ha obscurecido las reconstrucciones de las divergencias profundas de la historia de la vida. Se ha sugerido que parte de la dificultad de resolver el origen profundo de los eucariontes se podría deber a que el genoma de estos sería producto de una fusión entre los genomas de una bacteria y de un metanógeno (McInerney y Wilkinson, 2005; Martin y Muller, 1998) pero la aplicación de los métodos filogenéticos habituales a los datos de genomas de bacterias, arqueas y eucariontes no han podido producir historias inequívocas sobre las relación entre los tres dominios. Un nuevo método llamado reconstrucción condicionada (Lake y Rivera, 2004) podría arrojar luz sobre en esta y otras cuestiones en las que la transferencia horizontal puede haber obscurecido las relaciones entre taxa. El método se basa en un modelo markoviano de evolución genómica (análogo a los modelos de substitución de nucleótidos que hemos revisado) que supone que en la historia de los genomas se pueden ganar o perder estocásticamente los genes produciendo patrones de ausencia o presencia de genes ortólogos en los genomas muestreados. Los conflictos entre los patrones de ausencia y presencia de ortólogos son análogos a los conflictos debidos a las substituciones múltiples en los métodos tradicionales y por la misma razón son un problema que se puede resolver si el fenómeno se puede modelar adecuadamente y no ocurre a tasas excesivamente altas.

Una innovación de este método es que puede usarse para determinar si el conflicto entre diferentes reconstrucciones se puede resolver mediante el

planteamiento de relaciones cíclicas (topológicamente similares a las redes de la descomposición de biparticiones) en lugar de relaciones con una topología de árbol, por lo que la reconstrucción condicionada puede, en principio, detectar eventos de fusión genómica, representándolos como ciclos (McInereney y Wilkinson, 2005). Además, como sólo la remoción de los genomas que serían resultado de fusiones puede romper los ciclos, esto puede servir para identificar a los genomas quiméricos. Potencialmente, esta misma idea se podría extender para ayudar en la identificación de genes cuya presencia en determinado genoma se debe a transferencia horizontal.

El algoritmo eBURST

Los métodos tradicionales de reconstrucción filogenética y de agrupamiento de poblaciones parecen prestarse mal para la representación de eventos evolutivos recientes como los que caracterizan a la diversificación de poblaciones microbianas con genotipos muy similares pues intentan reconstruir sus relaciones sin un modelo realista de la manera en que surgen y se diversifican los complejos clonales y tampoco permiten identificar cuáles son los genotipos fundadores de las diferentes poblaciones El algoritmo eBURST (con una implementación disponible en http://eburst.mlst.net) permite extraer este tipo de información a partir de datos de tipificación por secuenciación multilocus (MLST por sus siglas en inglés) u otros tipos de datos multilocus (Feil *et al*., 2004). El algoritmo se basa en encontrar aislados que compartan entre sí un número mínimo de alelos de los genes convencionales en la tipificación multilocus. El número mínimo de alelos compartidos es definido por el usuario y diferentes grados de astringencia puden servir para analizar diferentes aspectos de la ecología de los patógenos estudiados. Cuando las reglas para definir a un grupo son lo suficientemente astringentes, los agrupamientos obtenidos con eBURST pueden ser legítimamente considerados complejos clonales y además es posible identificar cuál es el genotipo fundador (Feil *et al*., 2004).

El algoritmo eBURST ha sido utilizado exitosamente para describir los patrones poblacionales de *Neisseria meningitidis* y de *Campylobacter jejuni*, organismos con altas tasas de recombinación, resolviendo correctamente la mayor parte de los complejos clonales de meningococos en un fundador primario rodeado de descendientes variantes en un sólo locus, un patrón más fácil de interpretar que el que se obtendría representando este tipo de datos mediante un dendrograma.

Los métodos que acabamos de mencionar requieren de mayor escrutinio teórico y empírico. Los métodos basados en modelos explícitos generalmente funcionan bien cuando el modelo es el adecuado, pero el impacto de las violaciones a sus supuestos (especialmente en el caso de la reconstrucción condicionada) aún no han sido suficientemente explorados. No obstante, estos métodos tienen el potencial de convertirse en el "siguiente paso" en el desarrollo de herramientas para el estudio de la evolución y la ecología con datos moleculares, especialmente en la medida en que su popularización vaya mostrando cuáles son las prácticas correctas para su uso.

Bibliografía

Abascal, F., R. Zardoya y D. Posada. 2005. ProtTest: selection of best-fit model of protein evolution. *Bioinformatics* 21:2104-2105.

Adachi, J. y M. Hasegawa. 1995. MOLPHY: Programs for molecular phylogenetics. Institute of Statistical Mathematics. Tokio, Japón.

Adachi, J. y M. Hasegawa. 1996. Model of amino acid substitution in proteins encoded by mitochondrial DNA. *Journal of Molecular Evolution* 42: 459-468.

Almeida, J.S. y S. Vinga. 2002. Universal sequence map (USM) of arbitrary discrete sequences. *BMC Bioinformatics* 3:6.

Anderson, J. F., T. G. Andreadis, C. R.Vossbrinck, S. Tirrell, E. M. Wakem, R.A. French, A. E. Garmendia y H.J. Van Kruiningen. 1999. Isolation of West Nole virus form mosquitoes, crows and a Cooper´s hawk in Connecticut. *Science* 286:2331-2333.

Anisimova, M., J.P. Biewlawski y Z.H. Yang. 2002. Accuracy and power of Bayes prediction of aminoacid sites under positive selection. *Molecular Biology and Evolution* 19: 950-958.

Aris-Brosou, S. y Z. Yang. 2002. Effects of model of rate evolution on estimation of divergence dates with special reference to the metazone 18s ribosomal RNA phylogeny. *Systematic Zoology* 51:703-714.

Barton, G. J. 2001. Creation and analysis of protein multiple sequence alignments. En: *Bioinformatics: a practical guide to the analysis of genes and proteins*. A. D. Baxevanis y B. F. F. Ouelette [eds.]. Wiley-Interscience. Nueva York.

Barton, G. J. y J. E. Sternberg. 1987. A strategy for the rapid multiple alignment of protein sequences. Confidence levels from tertiary structure comparisons. *Journal of Molecular Biology* 198:327-337.

Blaisdell, B.E. 1986. A measure of the similarity of sets of sequences not requiring sequence alignment. *Proceedings of the National Academy of Sciences of the USA* 83: 5155-5159.

Brinkman, F. S. L. y D. L. Leipe. 2001. Phylogenetic analysis. En: *Bioinformatics: a practical guide to the analysis of genes and proteins*. A. D. Baxevanis y B. F. F. Ouelette [eds.] Wiley-Interscience. Nueva York.

Bush, R.M., C.A. Bender, K. Subbarao, N.J. Cox y W.M.Fitch. 1999. Predicting the evolution of human influenza A. *Science* 286:1921-1925.

Chang, B.S., K. Jonsson, M.A. Kazmi, M.J. Donoghue y T.P. Sakmar. 2002. Recreating a functional ancestral archosaur visual pigment. *Molecular Biology and Evolution* 19:1483-1489.

Chang, J. 1996. Inconsistency of evolutionary tree topology reconstruction methods when substitution rates vary across characters. *Molecular Biosciences* 134:189-215.

Chen, X., S. Kwong y M. Li. 1999. A compression algorithm for DNA sequences and its application in genome comparison. *Genome Informatics Series/Workshop on Genome Informatics* 10:51-61.

Christoffels, A., A. van Gelder, G. Greyling, R. Miller, T. Hide y W. Hide. 2001. STACK: Sequence tag alignment and consensus knowledgebase. *Nucleic Acids Research* 29:234-238.

Churchill, G.A., A. von Haesler y W.C. Navidi. 1992. Sample size for a phylogenetic inference. *Molecular Biology and Evolution* 9:753-769.

Dayhoff, M.O., R.M. Schwartz y B.C. Orcutt. 1978. A model of evolutionary change in proteins. En: M. O. Dayhoff [ed.]. *Atlas of protein sequence and structure*. National Biomedical Research Foundation. Washington.

Doolittle, R.F. 1986. *Of URFs and ORFs: a primer on how to analyze derived amino acid sequences*. University Science Books. Mill Valley, California.

Eddy, S.R. 1995. Multiple alignment using hidden Markov models. *Proceedings of the Third International Conference on Intelligent Systems for Molecular Biology*. 3:114-120.

———. 2004. What is dynamic programming? *Nature Biotechnology* 22:909-910.

Edgar, R.C. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research* 32:1792-1797.

Felsenstein, J. 1981. Evolutionary trees from DNA sequences: a maximum likelihood approach. *Journal of Molecular Evolution* 17:368-376.

———. 1999. PHYLIP. Phylogeny inference package. Version 3.572. University of Washington. Seattle. Disponible en: http://evolution.genetics.washington.edu/phylip.html.

———. 2001 Taking variation of evolutionary rates between sites into account in inferring phylogenies. *Journal of Molecular Evolution* 53:447-455.

Feng, D.F. y R.F. Doolittle. 1987. Progressive sequence alignment as a prerequisite to correct phylogenetic trees. *Journal of Molecular Evolution* 25:351-360.

Fitch, W.M. 1971. Rate of change of concomitantly variable codons. *Journal of Molecular Evolution* 1:84-96.

———. 1977. On the problem of finding the most parsimonious tree. *The American Naturalist* 111:223-257.

———. 2000. Homology: a personal view on some of the problems. *Trends in Genetics* 16:227-231.

Fitch, W.M. y E. Markowitz. 1970. An improved method for determining codon variability in a gene and its application to the rate of fixation of mutations in evolution. *Biochemical Genetics* 4:579-593.

Fleming, M. A., J.D. Potter, C.J. Ramírez, G.K. Ostrander y E.A. Ostrander. 2001. Understanding missense mutations in the BRCA1 gene: an evolutionary approach. *Proceedings of the National Academy of Sciences USA* 100:1151-1156.

Galtier, N. 2001. Maximum-likelihood phylogenetic analysis under a covarion-like model. *Molecular Biology and Evolution* 18:866-873.

Gatesy, J., R. DeSalle y W. Wheeler. 1993. Alignment-ambiguous nucleotide sites and the exclusion of systematic data. *Molecular Phylogentics and Evolution* 2:152-157.

Gibas, C. y P. Jambeck. 2001. *Developing bioinformatics computer skills*. O'Reilly. Sebastopol, California.

Giribit, G. 2001. Exploring the behavior of POY, a program for direct optimization of molecular data. *Cladistics* 17:S60-S70.

Gladstein, D. y W. Wheeler. 1996. POY: *Software for direct optimisation of DNA and other data.* American Museum of Natural History (ftp://ftp.amnh.org/pub/molecular/poy).

Goad, W.B. y M.I. Kanehisa. 1982. Pattern recognition in nucleic acid sequences I. A general method for finding local homologies and symmetries. *Nucleic Acids Research* 10:247-263.

Goldman, N. y S. Wheelan. 2000. Statistical tests of γ-distributed rate heterogeneity in models of sequence evolution in phylogenetics. *Molecular Biology and Evolution* 17:974-978.

Goldman, N. y Z. Yang. 1994. A codon-based model of nucleotide substitution for protein-coding DNA sequences. *Molecular Biology and Evolution* 11:725-736.

Goldman, N., J. P. Anderson y A.G. Rodrigo. 2000. Likelihood-based tests of topologies in phylogenetics. *Systematic Biology* 49:652-670.

Goloboff, P.A. 1998. *Principios básicos de cladística*. Sociedad Argentina de Botánica. Buenos Aires.

Gonnett, G.H., M.A.Cohen y S.A. Benner. 1992. Exhaustive matching of the entire protein sequence database. *Science* 256:1443-1145.

Gotoh, O. 1996. Significant improvement in accuracy of multiple protein sequence alignments by iterative refinement as assessed by reference to structural alignments. *Journal of Molecular Biology* 264:823-838.

Gupta, S.K., J.D. Kececioglu y A.A. Schaffer. 1995. Improving the practical space and time efficiency of the shortest-paths approach to sum-of-pairs multiple sequence alignment. *Journal of computational biology* 2:459-472.

Hasegawa, M., H. Kishino y T.-A. Yano. 1985. Dating the human-ape splitting by a molecular clock of mitochondrial DNA. *Journal of Molecular Evolution* 22:160-174.

Henikoff, S. y J.G. Henikoff. 1991. Automated assembly of protein blocks for database searching. *Nucleic Acids Research* 19:6565-6572.

Henikoff, S. y J.G. Henikoff. 1992. Amino acid substitution matrices from protein blocks. Proceedings of the National Academy of Sciences USA.89:10915-10919.

Hillis, D.M. 1994. Homology in molecular biology. En: B.K. Hall [ed.]. *Homology, the hierarchical basis of comparative biology*. Academic Press.

Huelsenbeck, J.P. y F.R. Ronquist. 2001. MrBayes: Bayesian estimation of phylogenetic trees. *Bioinformatics* 17:754-755.

Huelsenbeck, J.P. y K.A. Crandall. 1997. Phylogeny estimation and hypothesis testing using maximum likelihood. *Annual Review of Ecology and Systematics* 28:437-466.

Hughes, J.M., C.J. Peters, M.L. Cohen y B.W. Mahy. 1993. Hantavirus pulmonary syndrome: an emerging infectious disease. *Science* 262:850- 851.

Janecek, L.L., R.L. Honeycutt, R.M. Adkins y S.K. Davis. 1996. Mitochondrial gene sequences and the molecular systematics of the artiodactyl subfamily Bovinae. *Molecular Phylogenetics and Evolution* 6:107-119.

Janke, A., D. Erpenbeck, M. Nilsson y U. Arnason. 2001. The mitochondrial genomes of the iguana (Iguana iguana), and the caiman (Caiman crocodylus): implications for amniote phylogeny. *Proceedings of the Royal Society of London B Biological Sciences* 268:623-631.

Jeffrey, H.J. 1990. Chaos game representation of gene structure. *Nucleic Acids Research* 18:2163-2170.

Jones, D.T., W.R. Taylor y J.M. Thornton. 1992. The rapid generation of mutation data matrices form protein sequences. *Computer Applications in Biosciences* 8:275-282.

Jones, D.T., W.R. Taylor y J.M. Thornton. 1994. A mutation data matrix for transmembrane proteins. *FEBS Letters* 33:3038-3049

Jukes, T.H. y C.R. Cantor. 1969. Evolution of protein molecules. En: H.N. Munro [ed.]. *Mammalian Protein Metabolism.* Academic Press. Boca Raton, Florida.

Kimura, M. 1980. A simple method for estimating evolutionary rate of base substitution through comparative studies of nucleotide sequences. *Journal of Molecular Evolution* 16:111-120.

———. 1983. *The neutral theory of molecular evolution*. Cambridge University Press. Cambridge.

Kishino, H. y M. Hasegawa. 1989. Evaluation of the maximum-likelihood estimate of the evolutionary tree topologies from DNA sequence data and the branching order in hominoidea. *Journal of Molecular Evolution* 29:170-179.

Kolaczkowski, B. y J.W. Thornton. 2004. Performance of maximum parsimony and likelihood phylogenetics when evolution is heterogeneous. *Nature* 431:980-984.

Kuhner, M.K. y J. Felsenstein. 1994. A simulation comparison of phylogeny algorithms under equal and unequal evolutionary rates. *Molecular Biology and Evolution*. 11:459-468.

Kumar, S., K. Tamura, I.B. Jakobsen y M. Nei. 2001. MEGA 2: Molecular evolutionary analysis software. *Bioinformatics* 17:1244-1245.

Lake, J.A. y M.C. Rivera. 2004. Reconstructing evolutionary trees from DNA and protein sequences: paralinear distances. *Molecular Biology and Evolution* 21:681-690.

Lanciotti, R.S., J.T. Roehrig, V. Deubel, J. Smith, M. Parker, K. Steele, B. Crise, K.E. Volpe, M.B. Crabtree, J.H. Scherret, R.A. Hall, J.S. MacKenzie, C.B. Cropp, B. Panigrahy, E. Ostlund, B. Schmitt, M. Malkinson, C. Banet, J. Weissman, N. Komar, H.M. Savage, W. Stone, T. McNamara, D.J. Gubler. 1999. Origin of the West Nile virus responsible for an outbreak of encephalitis in the North eastern United States. *Science* 286:2333-2337.

Lee, M.S.Y. 2001. Unalignable sequences and molecular evolution. *Trends in Ecology and Evolution* 16: 681-685.

Lewis, P.O. 1998. Maximum likelihood as an alternative to parsimony for inferring phylogeny using nucleotide sequence data. En: D.E. Soltis, P.S. Soltis y J.J. Doyle [eds.]. *Molecular suystematics of plants II*. Kluwer. Boston, Massachusetts.

———. 2001a. A likelihood approach to estimating phylogeny from discrete morphological character data. *Systematic Biology* 50:913-925.

———. 2001b. Phylogenetic systematics turns over a new leaf. *Trends in ecology and evolution* 16:30-37.

Li, M., J.H. Badger, X. Chen, S. Kwong, P. Kearne y H. Zhang. 2001. An information-based sequence distance and its application to whole mitochondrial genome phylogeny. *Bioinformatics* 17:149-154.

Li, W.-H. 1997. *Molecular evolution*. Sinauer Associates. Sunderland, Massachussetts.

Li, W.-H. y D. Graur. 1991. *Fundamentals of molecular evolution*. Sinauer Associates. Sunderland, Massachussetts.

Lipman, D.J. y W.R. Pearson. 1985. Rapid and sensitive protein similarity searches. *Science* 227:1435-1441.

Lockhart, P.J., D. Huson, U. Maier, M.J. Fraunholz, Y. Van de Peer, A.C. Barbrook, C.J. Howe y M.A. Steel. 2000. How molecules evolve in bacteria. *Molecular Biology and Evolution* 17:835-838.

Lockhart, P.J., M.A. Steel, A.C. Barbrook, D. Huson, M.A. Charleston, C.J. Howe. 1998. A covariotide model explains apparent phylogenetic structure of oxygenic photosynthetic lineages. *Molecular Biology and Evolution* 15:1183-1188.

Lutzoni, F., P. Wagner, V. Reeb y S. Zoller. 2000. Integrating ambiguously aligned regions of DNA sequences in phylogentic analyses without violating positional homology. *Systematic Biology* 49:628-651.

McInerney, J.O. y M. Wilkinson. 2005. New methods ring changes for the tree of life. *Trends in Ecology and Evolution* 20:105-107.

Morgenstein, B., A. Dress y T. Werner. 1996. Multiple DNA and protein sequence alignment based on segment-to-segment comparison. *Proceedings of the National Academy of Sciences USA* 93:12098-12103.

Morrone, J.J. 2000. *El lenguaje de la cladística*. Universidad Nacional Autónoma de México. México, D.F.

Mouse Genome Sequencing Consortium. 2002. Initial sequencing and comparative analysis of the mouse genome. *Nature* 420:520-562.

Muse, S.V. 1995. Evolutionary analyses of DNA sequences subject to constraints on secondary structure. *Genetics* 139:1429-1439.

Muse, S.V. y B.S. Gaut. 1994. A likelihood approach for comparing synonymous and non-synonymous substitution rates with application to the chloroplast genome. *Molecular Biology and Evolution* 11:715-724.

Needleman, S.B. y C.D. Wunsch. 1970. A general method applicable to the search for similarities in the amino acid sequence of two proteins. *Journal of Molecular Biology* 48:443-453.

Nei, M. 1987. *Molecular evolutionary genetics*. Columbia University Press. Nueva York.

Nei, M. y S. Kumar. 2000. *Molecular Evolution and Phylogenetics*. Oxford University Press. Nueva York.

Nielsen, R. y J.P. Huelsenbeck. 2002. Detecting positively selected amino acid sites using posterior predictive p-values. En: R.B. Altman, A.K. Dunker, L. Hunter, K. Lauderdale y T.E. Klein, [eds.]. *Pacific Symposium on Biocomputing*. Proceedings. World Scientific, Singapore.

Notredame, C. 2002. Recent progress in multiple sdequence alignment: a survey. *Pharmacogenomics* 3:131-144.

Notredame, C. y D.G. Higgins. 1996. SAGA: sequence alignment by genetic algorithm. *Nucleic Acids Research* 24:1515-1524.

Notredame, C., D.G. Higgins y J. Heringa. 2000. T-COFFEE: a novel method for fast and accurate multiple sequence alignment. *Journal of Molecular Biology* 302:205-217.

Nylander, J.A.A. 2004. MrModelTest v2. Programa distribuido por el autor. Evolutionary Biology Center. Uppsala University. Uppsala.

Ota, R., P.J. Waddell, M. Hasegawa, H. Shimodaira y H. Kishino. 2000. Appropriate likelihood ratio tests and marginal distributions for evolutionary tree models with constraints on parameters. *Molecular Biology and Evolution* 17:798-803.

Page, R.D.M. y E.C. Holmes. 1998. *Molecular evolution a phylogenetic approach*. Blackwell Science. Oxford.

Penny, D., B.J. McComish, M.A. Charleston y M.D. Hendy. 2001. Mathematical elegance with biochemical realism: the covarion model of molecular evolution. *Journal of Molecular Evolution*. 53:711-723.

Petrilli, P. 1993.Classification of protein sequences by their dipeptide composition. *Computer Applications in the Biosciences* 9:205-209.

Posada, D. y K.A. Crandall. 1998. Modeltest: testing the model of DNA substitution. *Bioinformatics* 14:817-818.

Posada, D. y K.A. Crandall. 2001. Selecting the best-fit model of nucleotide substitution. *Systematic Biology* 50:580-601.

Rodríguez, F., J.L. Oliver, A. Marín y J.P. Medina. 1990. The general stochastic model of nucleotide substitution. *Journal of Theoretical Biology* 142:485-501.

Rota, P.A., M.S. Oberste, et al. 2003. Characterization of a novel coronavirus associated with severe acute respiratory syndrome. *Science* 5624:1394-1399.

Rzhetsky, A. y M. Nei. 1993. Theoretical foundations of the minimal-evolution method of phylogenetic inference. *Molecular Biology and Evolution*. 10:1073-1095.

Rzhetsky, A. y M. Nei. 1994. METREE: a program package for inferring and estimating minimum evolution trees. *Computer Applications in the Biosciences* 10:409-412.

Saitou, N. y M. Nei. 1987. The neighbor joining method: a new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution* 4:406-425.

Schuler, G.D. 2001. Sequence alignment and database searching. En: A.D. Baxevanis y B.F.F. Ouelette [eds.]. *Bioinformatics: a practical guide to the analysis of genes and proteins*. Wiley-Interscience. Nueva York.

Sennblad, B. y B. Bremer. 2000. Is there a justification for differential a priori weighting in coding sequences? A case study ffrom rbcL and Apocynaceae s.l. *Systematic Biology* 49:101-113.

Smith, R.F. y T.F. Smith. 1992. Pattern-induced multi-sequence alignment (PIMA) algorithm employing secondary structure-dependent gap penalties for use in comparative protein modelling. *Protein Engineering* 5:35-41.

Smith, R.F., M.S. Waterman y W.M. Fitch. 1981. Comparative biosequence metrics. *Journal of Molecular Evolution* 18:38-46.

Sourdis, J y C. Krimbas. Accuracy of phylogenetic trees estimated from DNA sequence data. *Molecular Biology and Evolution* 4:159-165.

Stanhope, M.J., O. Madsen,V.G. Waddell, G.C. Cleven, W.W. de Jong y M.S. Springer. 1998. Highly congruent molecular support for a diverse superordinal clade of endemic african mammals. *Molecular Phylogentics and Evolution* 9:501-508.

Stuart, G.W., K. Moffet y J.J. Leader. 2002b. A comprehensive vertebrate phylogeny using vector representations of protein sequences from whole genomes Molecular *Biology and Evolution* 19:554-562.

Stuart, G.W., K. Moffet y S. Baker. 2002a. Integrated gene and species phylogenies from unalignable whole genome protein sequences. *Bioinformatics* 18:100-108.

Suchard, M.A., R. Weiss y J.S. Sinsheimer. 2001. Bayesian selection of continuous-time Markov chain evolutionary models. *Molecular Biology and* Evolution 18:1001-1013.

Suchard, M.A., R.E. Weiss, K.S. Dorman y J.S. Sinsheimer. 2002. Oh brother where are thou? A Bayes factor test for recombination with uncertain heritage. *Systematic Biology* 51:715-738.

Swofford, D.L, G.J. Olsen, P.J. Waddell y D.M. Hillis. 1996. Phylogenetic inference. En: D.M. Hillis, C. Moritz y B.K. Mable [eds.]. *Molecular systematics*. Sinauer Associates. Sunderland, Massachusetts.

Swofford, D.L. 2000. PAUP*: Phylogenetic analisys using parsimony (*and other methods). Version 4.0b10. Sinauer. Sunderland, Massachusetts.

Taylor, W.R. 1998. Dynamic sequence databank searching with templates and multiple alignment. *Journal of Molecular Biology* 280:375-406.

Thompson, J.D., D.G. Higgins y T.J. Gibson. 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Research* 25:4876-4882.

Thompson, J.D., F. Plewniak y O. Poch. 1999. A comprehensive comparison of multiple sequence alignment programs. *Nucleic Acids Research* 27:2682-2690.

Thornton, J.W. y R. DeSalle. 2000. A new method to localize and test the significance of incongruence: detecting domain shuffling in the nuclear receptor superfamily. *Systematic Biology* 49:183-201.

Tillier, E.R. M y R.A. Collins. 1995. Neighbor-joining and maximum likelihood with RNA sequences: addressing the interdependence of sites. *Molecular Biology and Evolution* 12:7-15.

Tuffley, C. y M.A. Steel.1997. Modelling the covarion hypothesis of nucleotide substitution. *Mathematical BioSciences* 147:63-91.

Uzzel, T. y K.W. Corbin. 1971. Fitting discrete probability distributions to evolutionary events. *Science* 172:1089-1096.

Van de Peer, Y., S.A. Rensing, U.-G. Maier, R. DeWachter. 1996. Substitution rate calibration of small subunit ribosomal RNA identifies Chlorarachnida nucleomorphs as remnants of green algae. *Proceedings of the National Academy of Sciences USA* 93:7732-7736.

Vinga, S. y J. Almeida. 2003. Alignment-free sequence comparison - a review. Bioinformatics 19:513-523.

Waddell, P.J., D. Penny y T. Moore. 1997. Hadamard conjugations and modelling sequence evolution with unequal rates across sites. *Molecular Phylogenetics and Evolution* 8:33-50.

Waterman, M.S. 1984. General methods of sequence comparison. Mathematical Biology 46:473-500.

Waterman, M.S., J.J. Joyce y M. Eggert. 1991. Computer alignment of sequences. En: M.M. Miyamoto y J. Carcraft [eds.]. *Phylogenetic analysis of DNA sequences*. Oxford University Press, Nueva York.

Wheeler, D. 1986. The ins and outs of character analysis. *Systematic Zoology* 35:102-109.

Wheeler, W.C. 1995. Sequence alignment, parameter sensitivity, and the phylogenetic analysis of molecular data. *Systematic Biology* 44:321-331.

———. 1999. Fixed character states and the optimization of molecular sequence data. *Cladistics* 15:379-385.

Yang, Z. 1993. Maximum-likelihood estimation of phylogeny from DNA sequences when substitution rates differ over sites. *Molecular Biology and Evolution* 10:1396-1401.

———. 1994. Maximum-likelihood phylogenetic estimation from DNA sequences with variable rates over sites: approximate methods. *Journal of Molecular Evolution* 39:306-314.

———. 1997. PAML: a program package for phylogenetic analysis by maximum likelihood. *Computer Applications in Biosciences* 13:555-556.

———. 1996. Among-site rate variation and its impact on phylogenetic analyses. *Trends in Ecology and Evolution* 11:367-372.

Segunda parte

Temas fundamentales de la ecología molecular

Capítulo 5

Los análisis de paternidad, ¿para qué nos sirven?

Jordan Golubov y Jorge Ortega

La paternidad es una de las preguntas centrales en la ecología, pero muchas líneas de investigación en este tema no pudieron ser abordadas hasta hace poco menos de 30 años, dando recientemente origen a una gran industria en el humano relacionada con los casos legales de parentesco o estudios forenses (Avise, 1994; Hughes, 1998). Aunque la teoría estaba desarrollada desde hace ya unos años, la creación de sistemas de cómputo cada vez más poderosos ha dado el soporte necesario para que los análisis de paternidad sean posibles, y han ayudado a su vez a enriquecer la teoría. Por otro lado, el desarrollo de nuevos marcadores moleculares y sus tecnologías han hecho que los análisis de paternidad puedan realizarse a escalas razonables (en términos poblacionales) y con costos cada vez menores.

Desde el punto de vista ecológico, la información sobre paternidad es esencial para abordar temas como la identificación de cepas bacterianas en la epidemiología, en donde se ha usado ampliamente el *fingerprinting* (ya que la caracterización puede llegar a ser difícil; Welsh y McClelland, 1990), el análisis de la diversidad clonal (Li y Ge, 2001), el impacto de la endogamia en una población (veáse el capítulo 6 de este libro), la verificación de líneas de parentesco, determinación del tamaño efectivo de la población (véase el capítulo 3 de este libro) y estimación de la adecuación poblacional por vía masculina y femenina (Devlin y Ellstrand, 1990; Ellstrand y Marshall, 1985).

De cierta forma, la paternidad no puede ser totalmente determinada a través de observaciones conductuales directas, ya que en muchas ocasiones un individuo tiene más de una pareja durante la temporada reproductiva y es difícil identificar cuál de las parejas generó la progenie. En otros casos, como en tortugas marinas, no es posible identificar a los padres -que únicamente se encuentran en el mar- y sólo conocemos a las madres. Cuando las observaciones conductuales han mostrado no ser suficientes, los análisis de paternidad han generado resultados que no eran esperados dentro del sistema, como el éxito y la frecuencia de cópulas extramaritales en aves (Bekkum *et al.*, 2006; Vonhof *et al.*, 2006) o han confirmado lo previsto teóricamente como la adecuación inclusiva en hormigas (Kronauer *et al.*, 2006). Estos sistemas requerían de marcadores moleculares que dieran información detallada sobre la madre, los hijos y los padres putativos para resolver las incertidumbres sobre paternidad (Fleischer, 1996).

La paternidad puede ser determinada a partir de alelos que se heredan a las crías y que, si no están presentes en los otros individuos, éstos pueden ser excluidos como padres putativos. Para determinar la paternidad con mayor certeza es necesario utilizar una gran cantidad de loci, los cuales proveerán de suficiente información al sistema de caracterización genética de los individuos.

Existen diversos marcadores moleculares que se han utilizado para generar lo que comúnmente conocemos como *DNA fingerprinting* o huellas dactilares de ADN. Tradicionalmente estas "huellas digitales" se referían a los patrones característicos de un individuo, pero el término se ha utilizado también de manera reciente para la caracterización electroforética y cromatográfica de proteínas y más recientemente de moléculas de ADN (Anderson, 1974; Levin *et al.*, 1980). El término *DNA fingerprinting* fue inicialmente introducido por Jeffries (1985) para describir el patrón electroforético de los fragmentos de ADN semejando un código de barras que podía utilizarse como identificador de individuos. Posteriormente, se ha utilizado el término de "perfil de ADN" (*DNA profiling*) para describir el uso combinado de varios tipos de marcadores moleculares para la caracterización de individuos.

En este capítulo se aborda en primer lugar las bases teóricas y metodológicas que han surgido para el análisis de paternidad y en segundo término, algunas de las preguntas y ejemplos de su aplicación, principalmente de tipo biológico-ecológico.

La teoría y los métodos

En la identidad genética de los individuos intervienen dos factores: la herencia parental y la mutación novedosa. La mutación novedosa puede ser generada por diversas razones como la duplicación, desfasamiento en las cadenas de ADN, errores en la replicación y reparación del ADN, modificaciones químicas y mutación meiótica y/o mitótica de ADN repetido. En organismos diploides, la recombinación meiótica provoca un intercambio de genes de los progenitores de tal manera que los hijos son el resultado de una mezcla de genes de ambos padres. Por otro lado, los factores ambientales provocan cambios espontáneos del material genético (*i.e.* mutación), permitiendo su recombinación. Estas mutaciones ocurren aproximadamente cada 300 bases a lo largo del genoma humano (Vicard y Dawid, 2004; Chakraborty *et al.*, 1996), y son estables desde el punto de vista evolutivo, ya que se mantienen a lo largo del tiempo, lo cual permite utilizar los loci en los que se presentan para diferenciar individuos y poblaciones. Esta recombinación o mutación espontánea genera, al amplificar ciertas regiones del genoma con diversos marcadores moleculares, patrones de bandeo distintos. Así, la esencia de los análisis de paternidad radica en identificar las similitudes y diferencias en estos patrones de bandeo entre individuos de una población, y por lo tanto se basa en la variación genética que se encuentre a nivel de ADN.

Para realizar los análisis de paternidad se deben considerar dos aspectos: el tipo de marcador molecular que se va a utilizar y la forma en que se va a analizar mediante un modelo o una técnica computacional.

Para que un marcador molecular sea efectivo se debe considerar lo siguiente: (1) los marcadores utilizados no deben estar ligados; (2) cada marcador debe ser específico para un solo locus, de preferencia uno que esté caracterizado en el cromosoma; (3) el marcador debe detectar un número grande de alelos de manera constante y (4) se deben crear patrones de bandeo en geles que sean fácilmente interpretables. Los marcadores más utilizados son los microsatélites, que son una serie de repeticiones de un solo motivo compuesto de 1 a 6 bases (Hancock, 1998), debido a que son marcadores codominantes altamente polimórficos y densamente distribuidos en el genoma de los organismos eucariontes (Kashi y Soller, 1998) que generan mucha información y son lo suficientemente específicos para diferenciar individuos dentro de una población. También existen otros marcadores como RFLPs, RAPDs e ISSRs (véanse los capítulos 18 y 19 de este libro).

Estos marcadores revolucionaron el mundo de la ecología conductual y la genética de la conservación, debido a que es relativamente fácil realizar pruebas de paternidad en pequeñas muestras de tejido, ya que su genética permite hacer inferencias usando modelos genéticos explícitos basados en la observación de las frecuencias alélicas de determinados loci (Hancock, 1998). Actualmente los análisis con microsatélites se encuentran en una fase de crecimiento y se han aplicado a un gran número de especies (algunos referenciados en GENBANK). Recientemente se ha incorporado también el uso de secuencias de variación del ADN que ocurren en un solo nucleótido o mejor conocidas como SNPs (*single nucleotide polymorphism*), los microarreglos de ADN (*ADN microarrays*) y SAGE (*serial analysis of gene expression*).

Los métodos para analizar la variación en un marcador molecular comparten el principio de considerar que una gran proporción de la información genética es específica del individuo y difiere entre individuos, por lo que el análisis de esta fracción diferente es la que puede proporcionar una identidad específica a cada uno de ellos. Sin embargo, hay dos problemas que pueden afectar este análisis. El primero es cuando dos secuencias muestran más diferencias de las que se esperarían por mutación simple, especialmente cuando no se tienen estimaciones de las tasas de mutación de la secuencia que se está probando, lo que dificulta la asignación de la paternidad, ya que ninguno de los padres (aún cuando el padre real se encuentre dentro del grupo de padres putativos) tendría una buena probabilidad de ser asignado. El segundo problema es cuando hay coincidencia entre dos secuencias, y en estos casos la pregunta principal es si la coincidencia se debe a una mutación por azar o a un origen común. En los humanos éste último caso es casi imposible, ya que 2 genomas humanos típicos difieren en cerca de 3 millones de sitios y no hay dos personas que sean idénticas. Esta variación es tan obvia que los sistemas de detección sólo necesitan unos cuantos sitios de variación (*i.e.* HLA, MHC, etc.) para poder asignar la paternidad de manera relativamente confiable.

La manera más sencilla de establecer patrones de parentesco (o similitud) en la población es por medio de conteos directos de las frecuencias de las bandas generadas por los marcadores moleculares en una muestra aleatoria, pero representativa de la población. Este método es relativamente sencillo y es similar a otros índices utilizados en ecología que buscan la similitud entre dos individuos. La fórmula general para el coeficiente de Dice, denominado *D*, es:

$$D = 2 \bullet N_{ab} / (N_a + N_b)$$

donde N_{ab} es el número de bandas que es compartido por los dos carriles a y b, cada uno correspondiente a un individuo. N_a y N_b son el número de bandas presentes en cada carril, respectivamente. El valor de este índice, también denominado *S* –coeficiente de Nei y Li– se acota entre 0 y 1, donde 0 significa que no tienen bandas en común y 1 que todas las bandas se comparten.

Hay otros índices, como el coeficiente de Jaccard que toma la forma:

$$S_j = N_{ab} / N_a + N_b - N_{ab}$$

Ambos índices sólo toman en cuenta datos que se comparten y no, como con marcadores dominantes, la ausencia de bandas.

Aunque sencillo, este método puede realizarse sólo por pares de individuos y dado que es únicamente un índice de similitud, no tiene un valor estadístico, esto es, no se puede asignar una variación ni un grado de error a la estimación. La asignación de paternidad en estos casos se asigna a aquel individuo que comparta la mayor cantidad de bandas con el sujeto en consideración. El problema más claro al que nos enfrentamos es que si utilizamos varios marcadores podemos tener casos en los que no podemos distinguir entre los miembros de un grupo de padres; otro problema es que al aumentar la muestra, las comparaciones son cada vez más laboriosas.

Existe otro método que incorpora modelos desarrollados para genética de poblaciones. En éste se considera que dentro de la población no existe diferenciación en subpoblaciones, por lo que cada alelo se considera una muestra aleatoria e independiente del acervo genético de la población, y la frecuencia alélica de cada locus puede multiplicarse para calcular la frecuencia real en la población (Lewontin y Hartl, 1991; Chakraborty y Kidd, 1991). De esta manera podemos obtener una multitud de datos, que en teoría son independientes, que nos darían información acerca del parentesco. Un segundo avance es considerar las frecuencias alélicas dentro de cada población y acotar los padres putativos a aquellos individuos cuyas frecuencias alélicas se parezcan más a nuestra muestra.

Los métodos de asignación de paternidad que incluyen el cálculo de algún grado de error pueden dividirse en tres grandes grupos: 1) métodos de exclusión, 2) métodos de asignación fraccionada y categórica, y 3) métodos de reconstrucción genotípica. Todos ellos tratan de reconstruir patrones de parentesco entre los diferentes individuos de una población.

Exclusión

Se basan en las reglas de herencia mendeliana y utilizan las incompatibilidades entre parientes y progenie para rechazar hipótesis particulares de parentesco. Por ejemplo, si se considera un locus autosómico neutral con dos alelos *A* y *B* en una población bajo equilibrio de Hardy-Weinberg, esperaríamos una proporción de genotipos con el siguiente arreglo 25% *AA*, 50% *AB* y 25% *BB,* y por lo tanto una frecuencia alélica en donde $p = q$. Si una madre de genotipo *AA* tiene hijos de tipo *AA*, entonces todos los individuos dentro de la población con genotipo *BB* pueden ser excluidos como padres y la probabilidad de exclusión específica sería de 0.25 o 25%. La exclusión es un método que puede definir y asignar (en teoría) al par de padres de cada miembro de la progenie, especialmente en condiciones donde son pocos los padres potenciales y se tienen marcadores altamente polimórficos.

Un problema asociado a este tipo de análisis es la exclusión estricta, donde los errores en la tipificación de las bandas, los alelos nulos y los alelos falsos pueden llevar a exclusiones falsas. Si tenemos alelos falsos como artefacto, incrementa la posibilidad de exclusión de padres putativos por la presencia de alelos no compartidos. Además, este método puede ser impráctico cuando se tiene un gran número de padres, ya que el número de loci necesario para excluirlos incrementa y bajo la exclusión estricta, un solo error al comparar los alelos puede llevar a descartar un padre putativo.

Asignación fraccionada y categórica

Cuando la exclusión completa no es posible, podemos recurrir a diferentes métodos probabilísticos que puedan discriminar entre los candidatos de un grupo de datos. El método de la asignación fraccionada y categórica calcula las probabilidades (por máxima verosimilitud) de asignarle un progenitor a una progenie. Este método probabilístico se basa en el teorema de Bayes y determina la probabilidad de una hipótesis dados los datos que se tienen: *P* (Hipótesis | Datos). En términos generales el teorema de Bayes especifica lo siguiente:

$$P(\text{hipótesis} \mid \text{datos}) = P(\text{hipótesis})\, P(\text{datos} \mid \text{hipótesis}) / P(\text{datos})$$

La parte izquierda de la ecuación es lo que se llama la distribución de probabilidad posterior (la cantidad de interés). *P*(datos | hipótesis) es el

likelihood (máxima verosimilitud) que refleja la probabilidad de observar los datos dada la hipótesis. El denominador es simplemente una constante de normalización que refleja la probabilidad de los datos dados todos los posibles resultados.[1]

El método fraccionado se basa principalmente en calcular el logaritmo de la proporción de la máxima verosimilitud (*likelihood ratio* o *LOD*) para determinar probabilísticamente si un individuo es el padre de la progenie. El valor obtenido se divide entre la probabilidad de que el genotipo de la progenie dado un gene de un progenitor y el otro tomado de una muestra aleatoria de la población. En el caso en que las relaciones progenitor progenie sean de cero, entonces no se le asigna progenitor a la progenie. El valor obtenido se coteja con el valor de los otros posibles padres y el de valor más alto de LOD es considerado como el padre más probable de la progenie.

Por ejemplo, asumiendo que los loci no se encuentran ligados, tenemos los genotipos (2, 3 y 4) en un locus autosómico para tres individuos (*B* la progenie, *C* la madre por ejemplo y *D* el padre real entre muchos padres putativos), en donde conocemos el genotipo de la madre (*C*) pero no del padre (*D*).

El valor de LOD para que *D* sea el padre de *B* es:

$$\text{Valor LOD (D es padre de B)} = \log_e \frac{T\,(g_b \mid g_c, g_d)}{T'\,/g_{b'} \mid g_{c'})}$$

donde T (gb|gc, gd)" es la probabilidad de transición mendeliana (en forma de matriz) de gb sobre el valor de los loci de ambos padres (gc y gd) y T' (gb|gc) es la transición de gb dado que conocemos el valor de gc.

Por ejemplo, si tenemos un locus con dos alelos a_1 y a_2 con frecuencias p_1 y p_2 la matriz de transiciones T(gb|gc, gd) tiene la forma:

$$P\,(a_1a_1 \mid a_1a_1,a_1a_1)\;P\,(a_1a_1 \mid a_1a_1,a_1a_2)\;P\,(a_1a_1 \mid a_1a_2,a_1a_2)$$
$$P\,(a_1a_2 \mid a_1a_1,a_1a_2)\;P\,(a_1a_2 \mid a_1a_1,a_2a_2)\;P\,(a_1a_2 \mid a_1a_2,a_1a_2)$$

1 Las probabilidades condicionadas se calculan como: P(A|B) = P(A|B)/P(B). En términos sencillos estamos diciendo que la probabilidad del evento A dado el evento B, es el resultado de la probabilidad de un evento conjunto entre el espacio muestral en donde ha ocurrido el evento B. Para nuestro caso de paternidad, esto significa que P(A | B) es la probabilidad de que el hijo A sea producto del progenitor B, y sería la probabilidad de compartir alelos, entre todos los alelos de padres posibles (ver capítulo 4).

y *T(gb|gc, gd)* es la probabilidad de segregación de g_b dado un gen de g_c y tiene la forma

$$T(g_b \mid g_c) = \begin{matrix} P(a_1a_1 \mid a_1a_1)\, P(a_1a_1 \mid a_1a_2) \\ P(a_1a_2 \mid a_1a_1)\, P(a_1a_2 \mid a_1a_2) \end{matrix}$$

Tenemos como resultado las matrices

$$T(g_b \mid g_c, g_d) = \begin{matrix} 1 & 0.5 & 0.25 \\ 0.5 & 1 & 0.5 \end{matrix}$$

y

$$T(g_b \mid g_c) = \begin{matrix} p_1 & 0.5p_1 \\ p_2 & 0.5(p_1+p_2) \end{matrix}$$

La asignación fraccionada genera un valor acotado entre 0 y 1 a cada progenie de todos los padres no excluidos. Este mismo método puede ampliarse a muchos loci, de tal manera que los valores de LOD pueden sumarse para dar valores aditivos de LOD. La proporción de la progenie asignada a cada padre es proporcional a la verosimilitud de que sea el padre comparado con todos los demás padres no excluidos. Es decir, se utiliza la misma fórmula mencionada pero cambiando los valores asignados a los locus (ver detalles en Devlin *et al*., 1988). La diferencia con la exclusión completa es que no asignan el valor a un solo padre, sino que puede asignar la progenie a varios padres compatibles, lo cual incrementa el número de padres putativos para un individuo. Se podría considerar que la exclusión es biológicamente más real (un hijo solo puede tener un padre), pero el método fraccionado es estadísticamente más poderoso (Smouse y Meagher, 1994).

De estas dos rutas de desarrollo de los análisis de paternidad, la exclusión y la máxima verosimilitud (Adams *et al*., 1992; Milligan y McMurry, 1993), desafortunadamente ninguna ha generado resultados óptimos (Milligan y McMurry, 1993). La primera sólo es utilizada para determinar parentesco en poblaciones aisladas, pequeñas o controladas, porque necesita altos niveles de variación, entre 20 y 40 loci polimórficos (Chakraborty *et al*., 1988), mientras que la segunda no siempre puede asignar paternidad de manera adecuada, especialmente cuando la probabilidad de paternidad es igual entre dos o más individuos dentro de una misma población. Aunado a esto, se ha demostrado por simulación que estadísticamente la aproximación por máxima verosimili-

tud sobrerepresenta como posibles padres a individuos altamente homócigos (Devlin *et al*., 1988).

Reconstrucción genotípica parental

Este método utiliza los genotipos multilocus de un padre, que forzosamente se conoce *a priori*, y de la progenie para reconstruir los genotipos de los padres desconocidos que contribuyen al arreglo genómico de la progenie (Jones, 2001). La asociación de alelos entre loci provee de información acerca de los genotipos paternos que contribuyen al arreglo de la progenie.

Las técnicas actuales reconstruyen con algoritmos el número mínimo de genotipos paternos necesarios para explicar el juego de datos de la progenie. Para el caso en que la madre es conocida, todos los posibles genotipos paternos consistentes con al menos un hijo de la progenie son probados en combinación para determinar el número mínimo de padres que explican el arreglo completo de la progenie. Esta técnica es muy exigente en términos de cálculos en programas de computadora y no es consistente cuando hay arreglos de progenie con más de seis padres putativos (Jones *et al.*, 1998; Jones y Ardren, 2003).

También deben considerarse las dificultades asociadas al tipo de marcador (dominante y codominante), los de colecta (número y tipo de individuos colectados, sesgo en la colecta, etc.), y el programa utilizado, ya que algunos ignoran o dan un valor diferente a los alelos nulos, loci que se encuentren ligados, tipos de mutación, posibilidades de consanguinidad, entre otros (Smouse y Charkraborty, 1986). En la tabla 1 se muestran algunos de los programas que realizan análisis de paternidad y parentesco, sin olvidar que deben tomarse en cuenta los algoritmos que emplean y los marcadores para los cuales fueron diseñados.

Aplicación en animales

Los análisis de paternidad basados en marcadores moleculares en animales fueron inicialmente abordados por Jeffries y colaboradores (1985). Actualmente hay un auge de estos estudios (ver los últimos números de revistas como *Molecular Ecology*) debido a que permiten identificar patrones de cortejo y determinar la adecuación de los individuos, principalmente en sistemas sociales, donde algunos patrones conductuales no pueden ser observados directamente. Con estas herramientas se han encontrado evidencias de có-

pulas extra pareja en aves y mamíferos y se han desarrollado nuevas teorías sobre su efecto en la selección sexual (Fleischer, 1996); también se ha podido cuantificar y describir con certidumbre el sistema de apareamiento presente en una población (*e.g.* poliginia, poligamia, monogamia, etc.).

Asimismo, la identificación de la paternidad provee de información directa sobre el esfuerzo reproductivo, transmisión de atributos fenotípicos heredables (cálculo de heredabilidad), entre otros aspectos. Como ejemplo, en las poblaciones de macacos (*Macaca mulatta*), donde existe una fuerte dominancia jerárquica, se consideraba que el jefe de la tropa o macho alfa tenía ventajas en adecuación sobre los individuos de menores rangos. Sin embargo, Duvall *et al*. (1976) mostraron que sólo el 24% de la progenie producida durante un año era del macho alfa, por lo que la adecuación se encontraba distribuida entre varios miembros del grupo. Los análisis de paternidad identificaron así que en estos macacos el estatus social y la frecuencia de cópulas son malos indicadores del éxito reproductivo masculino.

Por otro lado, los análisis de parentesco han revolucionado la sociobiología de aves. Por ejemplo, se pensaba que los paseriformes eran aves monógamas por excelencia (Gill, 1990; Lack, 1968), pero se ha incrementado la evidencia de paternidad múltiple, en donde de 8 a 35% de los casos los padres putativos no son los padres reales, lo cual confirma que los sistemas de apareamiento en este grupo tienden a ser promiscuos (Griffith *et al*., 2002; Westneat y Stewart, 2003). Otro caso es el de las avestruces *Struthio camelus*, cuyos machos son territoriales. La evidencia molecular mostró que los machos son los padres de huevos de individuos en territorios vecinos, y esto se debe a que las hembras ponían sus huevos en nidos de hembras vecinas, distribuyendo así el riesgo de un ataque por depredadores (Kimwele y Graves, 2003).

En peces e insectos los análisis de paternidad se desarrollaron a mediados de los noventa; destacan los trabajos de detección de nuevas mutaciones en camadas grandes, el descubrimiento de nuevos sistemas de apareamiento a detalle y el estudio de formas de cuidado parental.

Hasta ahora, la evidencia molecular de parentesco se ha combinado con evidencia demográfica conductual, pero su relación con la conservación de las especies es más reciente (Gibbs *et al*., 1990; Kempenaers *et al*., 1992). Desde este punto de vista, los marcadores moleculares han servido para detectar cuellos de botella y otros importantes eventos históricos en una población, estimar el tamaño efectivo de las poblaciones, detectar procesos de selección (véase el capítulo 1 de este libro), determinar parentesco y sexo, estructura

de poblaciones, tasas de dispersión, procesos de migración, dietas, tamaños poblacionales, estatus de algunas enfermedades y detectar niveles de introgresión (véase el capítulo 8).

Aplicación en plantas

Las aplicaciones de los análisis de paternidad se encuentran íntimamente ligadas al sistema de apareamiento y la estructura clonal del organismo de estudio. Dada la común propagación vegetativa en plantas, los análisis de paternidad han ayudado a identificar y separar ramets de genets (ver capítulo 7). Esta separación es de suma importancia desde el punto de vista demográfico (la contabilidad de individuos en campo no es necesariamente el número de genotipos en el ambiente) y genético (en cuanto a la definición de unidades de selección y cuantificación de la adecuación), ya que se pueden tener errores demográficos con diversas implicaciones. Algunos trabajos han identificado genets dentro de una población (Graham *et al*., 1997), además de describir y correlacionar la identidad de cada individuo a nivel espacial (Maddox *et al*., 1989; Burke *et al*., 2000; Clark-Tapia, 2004).

Como se mencionó anteriormente, una de las aplicaciones más importantes de estos análisis es la estimación de la adecuación, que cuenta con dos componentes, el femenino y el masculino (Richards, 1997). El componente femenino comúnmente contempla la elección de polen adecuado para la fertilización y la producción de semillas y elementos relacionados con su dispersión (básicamente la producción de frutos). El componente masculino consiste en la producción de polen y la fertilización exitosa de óvulos del mismo genotipo (autogamia o geitonogamia) o de diferentes genotipos (xenogamia). En ambientes controlados y aun en ambientes naturales, los componentes femeninos son relativamente fáciles de cuantificar, ya que la producción de semillas da una idea general de la ganancia en adecuación por vía femenina. Por el contrario, el componente masculino es más difícil de seguir. Muchos estudios se enfocan en la producción de polen y algunos estiman su flujo por medio de marcadores fluorescentes. Sin embargo, la mayoría tiene supuestos que no siempre se cumplen en condiciones naturales (Campbell, 1991), y la remoción de polen no siempre puede ser correlacionada con el éxito de apareamiento (Ashman, 1998). La importancia de estudiar el componente masculino se vuelve evidente, cuando se considera el principal elemento que influye en la evolución de los despliegues florales (Bell, 1985; Bertin, 1988) y un importante componente del flujo génico que moldea la diversidad genética

entre y dentro de las poblaciones (Levin y Kerster, 1974) (véanse los capítulos 16 al 19 de este libro) .

Sin embargo, pocos estudios han usado la asignación de paternidad para describir cuantitativamente el flujo de polen, básicamente por dos razones: 1) las bases teóricas se desarrollaron hace varias décadas, pero las técnicas moleculares necesarias para asignar la paternidad se limitaban a la variación que se podía detectar con isoenzimas, y 2) para poder determinar con exactitud la paternidad, es necesaria una gran cantidad de información molecular, que incrementa los tamaños de muestra y causa que los estudios se limiten a unos cuantos individuos o a poblaciones muy pequeñas. Con los nuevos marcadores y técnicas moleculares (véanse los capítulos 16 al 19 de este libro), el análisis de una gran cantidad de alelos es posible, además de las nuevas herramientas teóricas (Luikart y England, 1999; Lewis y Snow, 1992), que incluyen caracteres masculinos en el análisis (Smouse *et al.*, 1999).

Aún con el desarrollo de más marcadores, que pueden obtenerse de manera rápida y relativamente barata (ver capítulo 18), las poblaciones naturales siguen siendo problemáticas, simplemente porque una proporción del flujo de polen (en algunos casos >60%) provienen de individuos que se encuentran fuera de la población muestreada (Streiff *et al.*, 1999; Dow y Ashley, 1996; Krauss, 1999) y sólo en algunos casos la paternidad puede asignarse con cierta precisión (Krauss, 1999).

Conclusión

El potencial de los análisis de paternidad es mucho más amplio de lo que se podría pensar. En principio, poder identificar el flujo de gametos dentro y entre poblaciones es uno de los principios básicos de la genética de poblaciones. Además, hay un claro impacto sobre los tamaños efectivos de la población si se conoce el número y el éxito reproductivo de cada uno de los padres dentro de la población. En términos conductuales, el desarrollo de estas técnicas ha ayudado a describir sistemas complicados de cortejo, conducta y estructura social en muchos vertebrados y hasta en invertebrados. Desde el punto de vista ético, el uso de marcadores moleculares para la conservación de las especies permite detener la extinción de las mismas y adecuar su manejo con bases genéticas firmes que no supongan un fracaso en cualquier acción establecida.

Tabla 1. Programas de cómputo para la aplicación de distintos métodos de análisis de paternidad y parentesco con marcadores moleculares. DC= diploide codominante, DD = diploide dominante, LS = loci ligados al sexo, C = loci citoplásmicos

Método	Programa	Marcador	Sitio
Exclusión	PROBMAX	DC, DD	http://www.uoguelph.ca/~rdanzman/software/PROBMAX/
	NEWPAT	DC, LS	http://www.zoo.cam.ac.uk/zoostaff/amos/newpat.html
	ALLELIX	DD	http://www.allelix.net
	ASPEX	DC	http://aspex.sourceforge.net
	GASSOC	DC	http://www.mayo.edu/statgen
	MERLIN	DC	http://www.sph.umich.edu/csg/abecasis/Merlin
	ERPA	DC	http://www.mds.qmw.ac.uk/statgen/dcurtis/software.html
	KINSHIP	DC	http://gsoft.smu.edu/
Asignación categórica	CERVUS	DC	http://helios.bto.ed.ac.uk/evolgen/cervus/cervus.html
	ACT	DD	http://www.epigenetic.org/Linkage/act.html
	BQTL	DC	http://hacuna.ucsd.edu/bqtl/
	FASTER	DC	http://www.hoschl.cz/faster/
	GENEHUNTER	DC	http://www.fhcrc.org/labs/kruglyak/Downloads/index.html
	PAPA	DC	http://www.bio.ulaval.ca/contenu-fra/professeurs/Prof-I-beratchez.html
	FAMOZ	DC, C, DD	http://www.pierroton.inra.fr/genetics/labo/Softeware/Famoz/
	PARENTE	DC	http://www2.ujf-grenoble.fr/leca/membres/manel.html
Asignación fraccionada	PATRI	DC	http://www.biom.cornell.edu/Homepages/Rasmus_Nielsen/files.html
Reconstrucción genotípica parental	GERUD	DC	http://www.biology.gatech.edu/professors/labsites/jones/parentage.html
	BAMA	DC	http://www.rni.helsinki.fi/~mjs/
	BLOCK	DC	http://www.cs.auc.dk/~claus/block.html
	FTM	DC	http://www.genealogy.com/soft_ftm.html
	GAP	DC	http://icarus2.hsc.usc.edu/epicenter/gap.html
	GRR	DC	http://bioinformatics.well.ox.ac.uk/GRR/

(Continúa)

Tabla 1. Programas de cómputo para la aplicación de distintos métodos de análisis de paternidad y parentesco con marcadores moleculares. DC= diploide codominante, DD = diploide dominante, LS = loci ligados al sexo, C = loci citoplásmicos *(continúa)*

Método	Programa	Marcador	Sitio
	MENDEL	DC	http://www.genetics.ucla.edu/software/mendel5
	PANGAEA	DC, DD	http://www.stat.washington.edu/thompson/Genepi/pangaea.html
	COPE	DC	http://www.infobiogen.fr/services/CoPE
	AMP	DC	http://watson.hgen.pitt.edu/biblio/Bibliography.html

Referencias de los programas de cómputo de la tabla 1

Act. De Andrade, M., C.I. Amos y T.J. Thiel. 1999. Methods to estimate genetic parameters for quantitative traits. *Genetic Epidemiology* 17:64-76.

Allelix . Para mayor información acudir a http://www.allelix.net.

Amp. Weeks, D.E. y K. Lange. 1988. The affected-pedigree-member method of linkage analysis. *American Journal of Human Genetics* 42:315-326.

Aspex. Hinds, D.A. y N. Risch. 1996. The ASPEX package: affected sib-pair exclusion mapping. http://aspex.sourceforge.net/.

Bama. Kilpikari, R. y M.J. Sillanpää. 2003. Bayesian analysis of multilocus association in quantitative and qualitative traits. *Genetic Epidemiology* 25:122-135.

Block. Para mayor información acudir a http://www.cs.auc.dk/~claus/block.html.

Bqtl. Borevitz, J.O., J.N. Maloof, J. Lutes, T. Dabi, J.L. Redfern, G.T. Trainer, J.D. Werner, T. Asami, C.C. Berry, D. Weigel y J. Chory. 2002. Quantitative Trait Loci Controlling Light and Hormone Response in Two Accessions of *Arabidopsis thaliana*. *Genetics* 160:683-96.

Cervus. Marshall, T.C., J. Slate, L.E.B. Kruuk y Pemberton. 1998. JM Statistical confidence for likelihood-based paternity inference in natural populations. *Molecular Ecology* 7: 639-655.

Cope. Brun-Samarcq, L., S. Gallina, A. Philippi, F. Demenais, G. Vaysseix y E. Barrillot . 1999. Cope: a collaborative pedigree drawing environment. *Bioinformatics* 15:345-346.

Delrious. Stone, J. y M. Björklund. 2001. Delrious: a computer program designed to analyse molecular marker data and calculate relatedness and delta estimates with confidence. *Molecular Ecology Notes* 1:209-212.

Erpa. Curtis, D. 1990. A program to draw pedigrees using LINKAGE or LINKSYS data files. *Annals of Human Genetics* 54:365-367.

Famoz. Gerber, S., P. Chabrier y A. Kremer. 2003. FAMOZ: a software for parentage análisis using dominant, codominant and uniparentally inherited markers. *Molecular Ecology Notes* 3: 479-481.

Faster. Para mayor información acudir a http://www.hoschl.cz/faster.

FTM. Family Tree ADN HISTORY UNEARTHED DAILY. Para mayor información acudir a http://www.genealogy.com/soft_ftm.html.

GAP. My Family Inc. Para mayor información acudir a http://icarus2.hsc.usc.edu/epicenter/gap.html.

Gassoc. Schaid, D.J. 1996. General score tests for association of genetic markers with disease using cases and their parents. *Genetic Epidemiology* 13:423–449.

Genhunter. Kong, A. y N.J. Cox. 1997. Allele-Sharing Models: LOD Scores and Accurate Linkage Tests. *American Journal of Human Genetics* 61:1179-1188.

Gerud. Jones, A.G. 2001. GERUD1.0: a computer program for the reconstruction multilocus ADN data. *Molecular Ecology Notes* 1:215–218.

Grr. Abecasis, G.R., S.S. Cherny, W.O. Cookson y L.R. Cardon. 2001. GRR: graphical of parental genotypes from progeny arrays using representation of relationship errors. *Bioinformatics* 17:742-3

Kinship. Singularity Solutions, Inc. Para mayor información acudir a http://gsoft.smu.edu.

Mendel. Lange, K., R. Cantor, S. Horvath, M. Perola, C. Sabatti, J. Sinsheimer y E. Sobel. 2001. Mendel version 4.0: A complete package for the exact genetic analysis of discrete traits in pedigree and population data sets. *American Journal of Human Genetics* 69(supplement): A1886.

Merlin. Abecasis, G.R., S.S. Cherny, W.O. Cookson y L.R. Cardon. 2002. Merlin-rapid analysis of dense genetic maps using sparse gene flow trees. *Nat Genet,* 30:97-101.

Newpat. Información: Amos, B. Department of Zoology, Cambridge, CB2 3EJ, UK. w.amos@zoo.cam.ac.uk.

Pangea. Información: Thompson, E., Department of Statistics, University of Washington, Seattle, WA 98195-4322, USA. thompson@stat.washington.edu

Papa. Dúchense, P., M.H. Godbout y L. Bernatchez. 2002. PAPA (Package for the Analysis of Parental Allocation) : A computer program for simulated and real parental allocation. *Molecular Ecology Notes* 2: 191-194.

Parente. Cercueil, A., E. Bellemain y S. Manel. 2002. PARENTE: computer program for parentage análisis. *Journal of Heredity* 93:458-459.PATRI. Nielsen, R., D.K. Mattila, P.J. Clapham y P.J. Palsbøll. 2001. Statistical approaches to paternity analysis in natural populations and applications to the North Atlantic humpback whale. *Genetics* 157:1673-1682.

Probmax. Danzmann, R.G. 1997. PROBMAX: A computer program for assigning unknown parentage in pedigree analysis from known genotypic pools of parents and progeny. *Journal of Heredity* 88:333.

Bibliografía

Adams, W. T., A.R. Griffin y G.F. Moran. 1992. Using paternity analysis to measure effective pollen dispersal in plant populations. *American Naturalist* 140:762-780.

Anderson, W.W. 1974. Frequent multiple insemination in a natural population of *Drosophila pseudoobscura*. *American Naturalist* 108:709-711.

Ashman, T.L. 1998. Is relative pollen production or removal a good predictor of male fitness? An experimental exploration with a wild strawberry (*Fragaria virginiana*). *American Journal of Botany* 85:1166-1171

Avise, J.C. 1994. *Molecular Markers, Natural History and Evolution*. Chapman & Hall, New York.

Bekkum, M.V., P.M. Sagar, J-C. Stahl y G.K. Chambers. 2006. Natal philopatry does not lead to population genetic differentiation in Bullers's albatross (*Thalassarche bulleri bulleri*). *Molecular Ecology* 15:73-79.

Bell, G. 1995. On the function of flowers. *Proceedings of the Royal Society Series B* 224:223-265.

Bertin, R.I. 1988. Paternity in plants. *En:* J. Lovett-Doust y L. Lovett-Doust [eds.]. *Plant reproductive ecology: patterns and strategies*. Oxford University Press.

Burke., J.M., M.R. Bulger, R.A. Wesselingh y M.L. Arnold. 2000. Frequency and spatial patterning of clonal reproduction in Louisiana Iris hybrid populations. *Evolution* 54:137-144.

Campbell, D.R. 1991. Comparing pollen dispersal and gene flow in a natural population. *Evolution* 45:1965-1968.

Chakraborty, R. y K. Kidd. 1991. The utility of ADN typing in forensic work. *Science* 254:1735-1739.

Chakraborty, R., T.R. Meagher y P.E. Smouse. 1988. Parentage analysis with genetic markers in natural populations. I. The expected proportion of offspring with unambiguous paternity. *Genetics* 118:527-536.

Chakraborty, R., D.N. Stivers, Z. Yixi. 1996. Estimation of mutation rates from parentage exclusion data: applications to STR and VNTR loci. *Mutation Research* 354:41-48.

Clark-Tapia, R. 2004. Estudio de la estructura clonal, demografía y biología reproductiva de *Stenocereus eruca*: cactacea endemica amenzada de las planicies de Magdalena, Baja California Sur. Tesis de Doctorado, Universidad Nacional Autónoma de México UNAM.

Devlin, B. y N.C. Ellstrand. 1990. The development and application of a refined method for estimating gene flow from angiosperm paternity analisis. *Evolution* 44:248-259.

Devlin, B., K. Roeder y N.C. Ellstrand. 1988. Fractional paternity assignment: theoretical and comparison to other methods. *Theoretical and Applied Genetics* 76:369-380.

Dow, B.D. y M.V. Ashley. 1996. Microsatellite analisis of seed dispersal and parentage of saplings in bur oak *Quercus macrocarpa*. *Molecular Ecology* 5:615-627.

Ellstrand, N.C. 1983. Multiple paternities within the fruits of the wild radish, *Rapahnus sativus*. *American Naturalist* 123:819-828.

Ellstrand, N.C. y D.L. Marshall. 1985. Interpopulation gene flow by pollen in wild radish. *Raphanus sativus*. *American Naturalist* 126:606-616.

Fleischer, R.C. 1996. Application of molecular methods to the assessment of genetic mating systems in vertebrates. *En:* J.D. Ferraris y S.R. Palumbi [eds.]. *Molecular zoology: advances, strategies and protocols*. Wiley-Liss, New York.

Gibbs, H.L., P.J. Weatherhead, P.T. Boag, B.N. White, L.M. Tabak y D.J. Hoysak. 1990. Realized reproductive success of polygynous red winged blackbirds revealed by ADN markers. *Science* 250:1394-1397.

Gill, F.B. 1990. *Ornithology*. W. H. Freeman & Co., New York.

Graham, J., G.R. Squire, B. Marshall y R.E. Harrison. 1997. Spatially dependent genetic diversity within and between colonies of wild raspberry (*Rubus idaeus*) detected using RAPD markers. *Molecular Ecology* 6:1001-1008.

Griffith, S.C., I.P.F. Owens y K.A. Thuman. 2002. Extra pair paternity in birds: a review as interspecific variation and adaptive function. *Molecular Ecology* 11:2195-2212.

Hancock, J.M. 1998. Microsatellite and other simple sequences: genomic context and mutational mechanisms. En: D.B. Goldstein y C. Schlötterer [eds.]. *Microsatellite evolution and aplications*. Oxford University Press, New York.

Hughes, C.R. 1998. Integrating molecular techniques with field methods in studies of social behavior: a revolution results. *Ecology* 79: 383-399.

Jeffries, A.J., V. Wilson y S.L. Tghein. 1985. Individual-specific "fingerprints" of human ADN. *Nature* 316:75-79.

Jones, A.G. 2001. GERUD 1.0.: A computer program for the reconstruction of parental genotypes from progeny arrays using multilocus ADN data. *Molecular Ecology Notes* 1:215-218.

Jones, A.G. y W.R. Ardren. 2003. Methods of parentage analysis in natural populations. *Molecular Ecology* 12:2511-2523.

Kashi, Y. y M. Soller. 1998. Funtional roles of microsatellites and minisatellites. En: D.B. Goldstein y C. Schlötterer [eds.]. *Microsatellite evolution and aplications*. Oxford University Press, New York.

Kaufman, S.R., P.E. Smouse y E. Alvarez-Buylla. 1998. Pollen mediated gene flow and differential male reproductive success in a tropical pioneer tree *Cecropia obtusifolia* Bertol. (Moraceae): a paternity analysis. *Heredity* 81:164-173.

Kempenaers, B., G.R. Verheyen, M. Van den Broeck, T. Burke, K. Van Broeckhoven y A.A. Dhondt. 1992. Extra-pair paternity results from female preferente for high quality males in the Blue Tit. *Nature* 357:494-496.

Kimwele, C.N. y J.A. Graves. 2003. A molecular genetic analysis of the communal nesting of the ostrich (*Struthio camelus*). *Molecular Ecology* 12:229-236.

Krauss, S.L. 1999. Complete exclusion of nonsires in an analysis if paternity in a natural population using amplified length polymophism AFLP. *Molecular Ecology* 8: 217-226.

Krawczak, M. y J. Schmidtke. 1998. ADN fingerprinting. BIOS Scientific Publishers Inc., Oxford.

Kronauer, D.J., C. Schoning y J. Boomsma. 2006. Male parentage in army ants. *Molecular Ecology* 15: 1147-1151.

Lack., D. 1968. *Ecological adaptations for breeding in birds*. Chapman & Hall, Londres.

Levin, D.A. 1975. Gametophytic selection in plants and animals. *En:* D.L. Mulcahy [ed.]. Gametophytic selection in *Phlox*. North Holland, Amsterdam.

Levin, D.A. y H.W. Kerster. 1974. Gene flow in seed plants. *Evolutionary Biology* 7:139-220.

Levin, D.A. 1981. Dispersal versus gene flow in plants. *Annals of the Missouri Botanical Garden* 68:233-253.

Levin, L., M. Asmussen y O. Olvera. 1980. Population genetics of Mexican *Drosophila*. V. A high rate of multiple insemination in a natural population of *Drosophila pseudoobscura*. *American Naturalist* 116:493-503.

Lewis, O.O. y A.A. Snow. 1992. Deterministic paternity exclusion using RAPD markers. *Molecular Ecology* 1:155-160.

Lewontin, R.C. y D. Hartl. 1991. Population genetics in forensic DNA typing. *Science* 254:1745-1750.

Li, A. y S. Ge. 2001. Genetic variation and clonal diversity of *Psammochloa villosa* (Poaceae) detected by ISSR markers. *Annals of Botany* 87:585-590.

Luikart, G. y England, P. 1999. Statistical analysis of microsatellite DNA data. *Trends in Ecology and Evolution* 14:253-256.

Maddox, G.D., R.E. Cook, P.H. Wimberger y S. Gardescu. 1989. Clone structure in four *Solidago altissima* (Asteraceae) populations: rhizome connections within genotypes. *American Journal of Botany* 76:318-326.

Meagher, T.R. 1986. Analysis of paternity within a natural population of *Chamaelirium luteum*. I. Identification of most-likely male parents. *American Naturalist* 128: 199-215.

Milligan, B.G. y C.K. McMurry. 1993. Dominant vs. codominant genetic markers in the estimation of male mating success. *Molecular Ecology* 2:275-283.

Richards, A.J. 1997. *Plant breeding systems*. Chapman & Hall, Londres.

Smouse, P.E. y R. Chakraborty. 1986. The use of restriction length polymorphisms in paternity analysis. *American Journal of Human Genetics* 38:918-939.

Smouse, P.E. y T.R. Meagher. 1994. Genetic analysis of male reproductive contributions in *Chamaelirium luteum* (L.) Gray (Liliaceae). *Genetics* 136:313-322.

Smouse, P.E., T.R. Meagher y C.J. Kobak. 1999. Parentage analysis in *Chamaelirium luteum* (L.) Gray (Liliaceae): Why do some males have higher reproductive contributions. *Journal of Evolutionary Biology* 12:1069-1077.

Streiff, R., A. Duciusso, C. Lexer, H. Steinkellner, J. Gloessl y A. Kremer. 1999. Pollen dispersal inferred from paternity analysis in a mixed oak stand of *Quercus robur* L. and *Q. petraea* (Matt) Liebl. *Molecular Ecology* 8:831-841.

Vicard, P. y A.P. Dawis. 2004. A statistical treatment of biases affecting the estimation of mutation rates. *Mutation Research* 547:19-33.

Vonhof, M.J., D. Barber, M.B. Fenton & C. Stobeck. 2006. A tale of two siblings: multiple paternity in big brown bats (*Eptesicus fuscus*) demonstrated using microsatellite markers. *Molecular Ecology* 15:241-247.

Welsh, J. y M. McClelland. 1990. Fingerprinting genomes using PCR with arbitrary primers. *Nucleic Acids Research* 18:7213-7218.

Westneat, D.F. y I.R.K. Stewart. 2003. Extra-pair paternity in birds: causes, correlates, and conflict. *Annual Review of Ecology and Systematics* 34:365-396.

Zietkiewicz, E., A. Rafalski y D. Labuda. 1994. Genome fingerprinting by simple sequence repeat (SSR) – anchored polymerase chaín reaction amplification. *Genomics* 20:176-183.

Capítulo 6

La estimación de la endogamia y la relación entre la tasa de fecundación cruzada y los sistemas reproductivos en plantas con flores: una interpretación de su significado evolutivo

César A. Abarca García y Adriana López Villalobos

La producción de genotipos variables es uno de los efectos biológicos más claros y generales de la evolución del sexo en eucariontes. (Hickey, 2000). Sin embargo, algo asombroso de ésta generalidad es que en muchos grupos de plantas y animales multicelulares la evolución del sexo está ligada a una combinación compleja de mecanismos reproductivos que impactan la estructura y evolución de las especies, particularmente por la forma en que permiten que recombinen los genes dentro del genoma (Holsinger, 2000).

Un claro ejemplo de la complejidad de los mecanismos de reproducción son los sistemas reproductivos y de apareamiento que han surgido independientemente tanto en plantas como en animales (Reynolds, 1996; Sakai y Weller, 1999; Webb, 1999). Tradicionalmente el sistema de apareamiento se define como los caracteres reproductivos que determinan los patrones de apareamiento en los organismos. Es decir, quién se aparea con quién y cómo lo hacen (Barrett y Eckert, 1990; Barrett *et al.*, 1996; Vogler y Kalisz, 2001). Los sistemas reproductivos, en cambio, se refieren a la forma en que se encuentran acomodadas espacial y temporalmente las estructuras reproductivas masculinas y femeninas dentro de los individuos en una población (Darwin, 1887). Evidentemente los sistemas reproductivos tienen consecuencias directas sobre el sistema de apareamiento.

La diversificación en forma y función de las flores está asociada con una amplia variedad de estrategias de apareamiento y sistemas reproductivos.

En las plantas la reproducción involucra el proceso previo y posterior a la polinización, que se traduce en la fertilización de óvulos. Las plantas influyen en su reproducción a través de rasgos florales que afectan el movimiento de polen dentro y entre sus propias flores, así como entre flores de otras plantas y a través de mecanismos fisiológicos que gobiernan el destino del polen una vez que éste llega al estigma. La mayoría de las especies presentan flores perfectas, es decir, las estructuras que producen el polen y los óvulos, están juntas en una sola flor (hermafroditas), pero también han evolucionado estrategias en donde las funciones reproductivas (femenina y masculina) están separadas en diferentes estructuras en la misma planta (monoicas) o bien completamente separadas en diferentes plantas (dioicas). Existen estrategias asociadas a los sistemas reproductivos que promueven la fecundación cruzada, disminuyen el desperdicio del polen y evitan los efectos negativos de la endogamia; como la separación temporal (dicogamia) y espacial (hercogamia) de las estructuras reproductivas y los sistemas de incompatibilidad que impiden la autofertilización.

Los múltiples diseños florales influyen de manera directa en la frecuencia con la que se unen los gametos de los individuos no emparentados para formar un cigoto. Por ejemplo, las flores hermafroditas autocompatibles con poca hercogamia aumentan sus probabilidades de autofecundación y por lo tanto de endogamia, pues aumentan la interferencia polen-estigma. En cambio, la fecundación cruzada se promueve con diseños florales especializados (como las especies con polimorfismo estilar o heterostilia) y/o cuando la cantidad de néctar y polen que atraen a los polinizadores aumenta. Por tanto, dichos atributos florales repercuten directamente sobre la distribución de la diversidad genética de las poblaciones, pues reducen o aumentan la endogamia (Holsinger, 2000).

Diversos estudios empíricos han mostrado, por ejemplo, que las plantas asexuales o que se autofecundan presentan menos alelos por locus, tienen menores niveles de diversidad genética dentro de sus poblaciones y que éstas están más diferenciadas entre sí que especies con fecundación cruzada (Schoen y Brown, 1991; Charlesworth y Charlesworth, 1995; Awadalla y Ritland, 1997; Charlesworth y Yang, 1998; Dvorak *et al.*, 1998; Kraft *et al.*, 1998; Liu *et al.*, 1999; Savolainen *et al.*, 2000; Baudry *et al.*, 2001; Wright *et al.*, 2002). Es por ello que las consecuencias evolutivas de promover la fecundación cruzada y el mantenimiento de altos niveles de variación genética, ha sido un tema ampliamente debatido en la literatura (Fisher, 1930; Muller, 1932; Lewontin, 1971; Williams, 1975; Maynard Smith, 1978; Crow, 1988; Charlesworth,

1993). La evidencia teórica indica la importancia de mantener la variabilidad genética, particularmente porque aumenta las probabilidades de respuestas adaptativas (Holsinger, 2000), sin embargo, la evidencia empírica aún no es concluyente al respecto (Hickey, 2000). Conocer en qué medida los sistemas reproductivos en las angiospermas promueven la fecundación cruzada es importante para aportar elementos a la discusión de sus consecuencias en términos evolutivos.

En este capítulo se revisan los principales métodos que estiman la endogamia en una población, con énfasis en la tasa de fecundación cruzada, particularmente a través de marcadores moleculares co-dominantes y dominantes. Posteriormente se compara la tasa de fecundación cruzada en diferentes sistemas reproductivos en las angiospermas. Finalmente se discute cómo la tasa de fecundación cruzada es afectada por el sistema reproductivo y sus principales consecuencias evolutivas.

Las distintas formas de medir la endogamia

El término endogamia es usado para describir varios fenómenos relacionados, todos ellos referentes a situaciones en donde los apareamientos ocurren entre parientes (endogamia biparental) o consigo mismos (endogamia uniparental) y que tienen como consecuencia que dos alelos homólogos heredados a la progenie sean idénticos entre sí por descendencia de un alelo ancestral en común (Waser, 1993). Es decir, los eventos que tienen como consecuencia apareamientos endogámicos, incrementan los niveles de homocigosis en una población (Hartl, 1980, 2000; Hedrick, 2000).

La endogamia puede ser estimada directamente por medio de un *pedigree* o árbol genealógico, particularmente cuando tenemos información de los miembros fundadores de una población y de su descendencia, o indirectamente por medio de un marcador morfológico o molecular.

Estimación de la endogamia por medio de pedigrees

La información derivada de un pedigree se usa para determinar el coeficiente de endogamia *F*, definido como la probabilidad de que dos alelos homólogos sean idénticos por descendencia (*ipd*), esto es, que ambos sean copias de un alelo ancestral común. En un individuo, esta probabilidad de *ipd* depende de la relación de ancestría que poseían sus progenitores (Hedrick, 2000). Por ejemplo, si un individuo heterócigo A_1A_2 se reproduce por autofecundación,

tiene un 50% de posibilidades de heredar 2 alelos ipd, pues tiene la probabilidad que ¼ de su progenie sea A_1A_1 y ¼ A_2A_2; de esa forma, ambas probabilidades suman 0.5. El otro 50% corresponde a progenie heterócigа. Ahora veamos cómo se calcula el pedigree de un individuo endogámico producido en la segunda generación.

Cuando tenemos un pedigree como el de la Figura 1, en el que dos primos hermanos *M* y *N* que poseen un abuelo *A* heterócigo (A_1A_2) producen un hijo endogámico *L*, podemos calcular el coeficiente de endogamia *F* de ese individuo *L*. Para ello, primero se omiten todos los individuos no emparentados, pues no contribuyen en nada a la endogamia. Después y sabiendo que el abuelo poseía un genotipo heterócigo, se calcula la probabilidad de que siendo *L* A_1A_1 ó A_2A_2, estos alelos sean *ipd*. Esta probabilidad Pr es igual a la probabilidad de que el alelo A_1 o A_2 pasen por tres generaciones a través de los dos progenitores. Es decir, para el primer caso en el que *L* es A_1A_1:

la Pr de que A_1 pase por medio de *M* = (1/2) (1/2) (1/2) = 1/8
y la Pr de que A_1 pase por medio de *N* = (1/2) (1/2) (1/2) = 1/8

donde 1/2 representa la probabilidad de segregación de un alelo entre generaciones. Debido a que la probabilidad de que A_1 pase por descendencia hasta *L* depende de ambas probabilidades, la probabilidad de A_1A_1 en *L* = (la probabilidad de que A_1 pase por medio de *M*) (la probabilidad de que A_1 pase por medio de *N*) = 1/64.

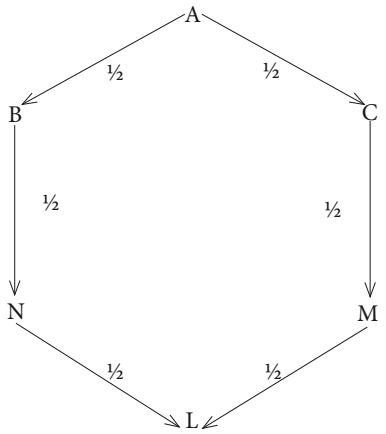
Del mismo modo, en caso de que L sea A_2A_2 tiene la misma probabilidad y se calcula de la misma forma, por lo que el coeficiente de endogamia del individuo L ó la probabilidad de *ipd* promedio es:

F = Pr de A_1A_1 + Pr de A_2A_2 = 1/32

El coeficiente de endogamia a partir de los cambios en la variabilidad genética

Debido a que la mayoría de las veces no tenemos la genealogía de los individuos y no sabemos cuánta endogamia había en las generaciones anteriores, a Wright (1965) se le ocurrió que podíamos asumir que en la población ancestral todos los alelos eran únicos. Es decir, que no había endogamia en esa población (*F*=0).

Figura 1. Pedigree de un individuo endogámico L, producto del apareamiento de dos primos hermanos N y M con un ancestro A heterócigo. ½ se refiere a la probabilidad de heredar un alelo idéntico por descendencia



De esta forma y sabiendo que con apareamientos endogámicos aumenta la tasa a la que se producen genotipos homócigos, pensó que la endogamia que había ocurrido entre generaciones se podía estimar calculando el descenso relativo en los niveles de variación genética que no deberían ocurrir en una población en donde los apareamientos fueran al azar. Para medir esta desviación en la frecuencia de heterócigos en una población, Wright desarrolló lo que se conoce como el coeficiente de endogamia (f_{is}), definido como:

$$f_{is} = 1\text{-}(H_o/H_e)$$

En donde H_o es la heterocigosis observada y H_e es la heterocigosis que esperamos cuando los apareamientos se dan al azar. Este coeficiente puede tomar un valor de –1 a 1. Si es de 0 no hay endogamia y la población está en equilibrio de Hardy Weinberg. Cuando el valor es negativo, significa que hay un exceso de heterócigos y si es positivo que hay un exceso de homócigos, ya que cuando $H_0 < He$ se asume que en esa población se produjeron más apareamientos endogámicos de lo que se espera por azar.

La tasa de fecundación cruzada con marcadores morfológicos

En muchos organismos hermafroditas, donde cada individuo puede producir una gran cantidad de hijos, como en la mayoría de las angiospermas, la progenie proviene de la combinación de la fecundación cruzada y de la autofecundación. La tasa a la que se producen un tipo u otro de cruzas tiene un significado evolutivo importante. Por ejemplo, teóricamente una tasa de autofecundación alta puede ser seleccionada en contra cuando incrementa la frecuencia de homócigos. De manera que muchos genes recesivos con efectos deletéreos se expresan y disminuye la adecuación de los individuos producto de la autofecundación. En esas circunstancias, cualquier mecanismo que evitara o disminuyera la endogamia y sus efectos negativos, sería favorecido por la selección natural (Barrett y Eckert, 1990). Sin embargo, la autofecundación también puede favorecerse en condiciones ecológicas particulares como cuando hay baja densidad de polinizadores, recurrentes cuellos de botella o efecto fundador (Loyd y Webb, 1992).

Los primeros modelos y estimaciones de la tasa de fecundación cruzada utilizaron caracteres morfológicos, ya que algunas especies presentan dos variedades morfológicas en una misma población que segregan de manera mendeliana, lo que puede ser utilizado para estimar la tasa de fecundación cruzada.

Datura stramonium es uno de esos casos raros en la naturaleza que tienen una característica invaluable, que el color del tallo segrega mendelianamente (Cuevas, 1996). Los tallos rojos corresponden al genotipo homócigo dominante y al heterócigo, mientras que los tallos verdes al genotipo homócigo recesivo. Con esa información se calculó la tasa de polinización cruzada denominada *t* para una población natural del estado de Morelos, siendo

$$t = H/p$$

donde *H* es la fracción de individuos heterócigos observados en la progenie de una planta de tallo verde (homóciga recesiva) y *p* la frecuencia del alelo marcador que se estimó a partir de la población parental, que en este caso es el alelo dominante que determina el tallo rojo. *H* fue calculada dividiendo el número de plántulas rojas, entre el total de plántulas germinadas por fruto (Allard y Workman, 1963).

La tasa promedio de polinización cruzada de la población fue de 17.1%, la más alta reportada para esta especie (esto quiere decir que 17 individuos

sobre 100 son producto de la cruza entre genotipos parentales diferentes). Esta variación puede deberse a factores genéticos o ecológicos que influyen directamente sobre el tipo de apareamientos en una población. En el caso de la población de *D. stramonium* de Morelos, una *t* promedio más grande que en otras localidades de Estados Unidos, parece estar correlacionada con una mayor actividad y eficiencia de sus polinizadores (esfíngidos).

La tasa de fecundación cruzada con marcadores moleculares co-dominantes: una revisión de los modelos desarrollados en las angiospermas

La segregación alélica de uno o varios loci en la progenie de individuos hermafroditas puede deberse, como ya se mencionó, a la combinación de autofecundación y fecundación cruzada, a partir de la cual se desarrolló el denominado modelo de apareamiento mixto (Hayman, 1953). Sin embargo, antes de analizar este modelo revisaremos el modelo de apareamientos al azar y luego el modelo de autofecundación completa.

En condiciones silvestres es difícil saber quién o quiénes fueron los donadores de polen de los frutos que observamos en una planta, pero podemos extraer su ADN y el de su progenie y reconstruir los eventos de apareamiento que dieron origen a tales frutos gracias a un marcador molecular co-dominante como las isoenzimas, o dominante como los RAPDs (véase capítulo 18 en esta obra).

Modelo de apareamientos al azar

Si los apareamientos en una población se dan al azar, se espera que la distribución genotípica de la progenie se comporte en términos del modelo de Hardy-Weinberg. Supongamos que tenemos los tres genotipos maternos posibles para un locus con dos alelos, el heterócigo A_1A_2 y los dos homócigos A_1A_1 y A_2A_2. Con apareamientos al azar esperamos una distribución de los tres genotipos de la progenie como la mostrada en la Tabla 1, en donde p y q representan la frecuencia de los alelos A_1 y A_2 en el polen de la población.

Las casillas izquierda inferior y derecha superior se encuentran vacías, porque en un evento de fecundación cruzada es imposible que una madre A_1A_1 y A_2A_2 produzcan hijos A_2A_2 y A_1A_1, respectivamente. Si los apareamientos son aleatorios, la distribución genotípica de la progenie dependerá de la frecuencia de cada uno de los dos alelos en el polen (A_1 y A_2) y del genotipo materno.

Tabla 1. Proporción genotípica esperada en la progenie en el modelo de apareamientos al azar

Genotipo materno	Genotipos en la progenie		
	A_1A_1	A_1A_2	A_2A_2
A_1A_1	p	q	-----
A_1A_2	½ p	½ (p + q)	½ q
A_2A_2	-----	p	q

Aunque en muchas poblaciones naturales es difícil esperar estas frecuencias genotípicas debido a que, por ejemplo, muchos fenotipos se aparean con fenotipos iguales (apareamientos asortativos positivos), el modelo sí constituye una valiosa hipótesis nula (Ritland y Jain, 1981; Ritland, 1983).

Modelo de apareamientos con autofecundación completa

La forma más extrema de endogamia en organismos que se reproducen sexualmente es la autofertilización. En estos casos, la tasa de fecundación cruzada denominada t es igual a 0 y la de autofecundación que podemos denominar s sería igual a s = 1 – t, es decir s = 1.

En este modelo (Ritland, 1985; Weir, 1996) consideremos un locus diploide que segrega para dos alelos A_1 y A_2, con una frecuencia de p y q en el polen. Con autofertilización esperaríamos una proporción genotípica de la progenie como la de la Tabla 2, donde H_o, P_o y Q_o representan las frecuencias genotípicas de A_1A_2, A_1A_1 y A_2A_2 respectivamente.

Tabla 2. Proporción genotípica esperada en el modelo con completa autofertilización

Genotipo materno	Genotipo de la progenie		
	A_1A_1	A_1A_2	A_2A_2
A_1A_1	p_o	-----	-----
A_1A_2	¼ H_o	½ H_o	½ H_o
A_2A_2	-----	-----	qo

La principal diferencia entre ambos modelos es que en el de apareamientos aleatorios las casillas intermedias representan la frecuencia de los dos alelos en el polen que donarían cada uno de los tres genotipos, y en éste, las casillas intermedias representan a los genotipos esperados después de la autofecundación.

Modelo de apareamientos mixtos

Éste modelo resulta de una combinación de los dos anteriores, es decir, del de autofecundación (s) y del de fecundación cruzada (t) (Ritland, 1983). En éste caso también podemos calcular la proporción de cada uno de los tres genotipos de la progenie a partir de una madre con un genotipo heterócigo A_1A_2 u homócigo A_1A_1 y A_2A_2. Como puede apreciarse en la Tabla 3, p y q representan la frecuencia de los alelos A_1 y A_2 en el polen, y t y s la tasa de fecundación cruzada y de autofecundación, respectivamente.

Tabla 3. Distribución genotípica de la progenie en el modelo de apareamientos mixtos

Genotipo materno	Genotipo de la progenie			
	A_1A_1	A_1A_2	A_2A_2	Otro genotipo
A_1A_1	s + tp	t q	-----	-----
A_1A_2	¼ s + ½ t	½ t (p +q)	¼ s + ½ t	t (1 – p – q)
A_2A_2	-----	t p	s + t q	-----

En éste modelo, al igual que en los anteriores, las casillas vacías corresponden a los genotipos que no se pueden producir.

A partir de este modelo podemos obtener una estimación directa de la tasa de fecundación cruzada t, si consideramos por ejemplo, que la proporción esperada de progenie heterógiga con una madre homóciga A_1A_1 es:

$$H = t\,q$$

pues de esta expresión podemos despejar t como:

$$t = H / q$$

y si tenemos que N_{12} es la progenie heteróciga del total de la progenie (N) segregada, entonces la estimación de la tasa de fecundación cruzada es:

$$t = N_{12} / q\, N$$

y como s + t =1, entonces también podremos saber qué proporción de la progenie proviene de autofecundación, simplemente como s = 1 – t.

Cuando tenemos marcadores moleculares muy variables como los microsatélites (ver capítulo 18), entonces las madres heterócigas son muy informativas. Es posible hacer una estimación de la tasa de fecundación cruzada usando toda la progenie heteróciga, excepto la A_1A_2 (parte derecha de la Tabla 3):

$$t = H / 1 - p - q$$

donde H es la progenie heteróciga distinta al genotipo materno.

Sin embargo, cuando estimamos únicamente la tasa de fecundación cruzada y autofecundación, podemos estar sobre-estimando la última, debido a que una parte de los apareamientos endogámicos se pueden dar entre parientes (endogamia biparental).

Generalmente, cuando hay endogamia biparental en una población existe estructuración espacial en forma de parches conformados por individuos emparentados que se cruzan entre sí. Si hay endogamia biparental, la t_s estará sesgada a una parte de la población. Por lo tanto, si calculamos la diferencia que existe entre la tasa de fecundación cruzada calculada a partir de múltiples loci o mutilocus (t_m) y la tasa calculada a partir del promedio de un solo locus o unilocus (t_s), tendremos una medida de la endogamia biparental.

Cuando hay endogamia biparental en una población la diferencia entre $t_m - t_s$ siempre positiva, esto es $t_m > t_s$, pues t_s estará sesgada o restringida a una parte de la población.

La tasa de fecundación cruzada a partir de marcadores dominantes

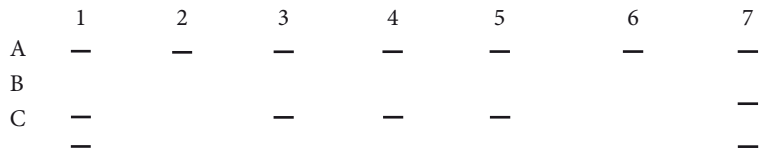
La tasa de fecundación cruzada en plantas ha sido tradicionalmente obtenida con isoenzimas. Sin embargo, en ausencia de suficiente polimorfismo, pueden emplearse distintos marcadores dominantes obtenidos con PCR, que generen de preferencia un alelo dominante (presencia de banda) en baja frecuencia en la población. Antes de continuar, hay que recordar brevemente que cuando utilizamos marcadores dominantes se asume que la presencia/ausencia

de bandas corresponde al fenotipo de un locus con dos alelos, en el que la presencia de la banda indica un genotipo homócigo dominante o heterócigo, mientras que la ausencia indica un genotipo homócigo recesivo (Williams, 1990; ver el capítulo 18).

Imaginemos que obtenemos el fenotipo de siete individuos de una misma familia, pero que no conocemos el genotipo materno, justo como se muestra en la Fig. 2. En esas circunstancias un alelo dominante (presencia de banda) que se encuentre en alta frecuencia en la población será el marcador menos informativo. La presencia de la banda A en todos los hijos no nos permite saber si la planta materna es homóciga recesiva, homóciga dominante o heteróciga y por lo tanto, tampoco podremos discriminar qué hijos provienen de un evento de fecundación cruzada. Es decir, esta configuración no es informativa para detectar los eventos de fecundación cruzada.

Por su parte, la ausencia de la banda B en pocos hijos sugiere que el genotipo materno podría ser heterócigo u homócigo recesivo y aunque esta configuración de bandas es más informativa, los eventos de fecundación cruzada tampoco podrán ser discriminados con certeza si nos basamos sólo en este marcador. Finalmente, la configuración más informativa es cuando el marcador está presente en pocos hijos (banda C). En este caso podemos inferir que la planta materna tiene un genotipo homócigo recesivo y la progenie que tiene el marcador es heteróciga y producto de fecundación cruzada.

Figura 2. Representación hipotética de la segregación de un marcador dominante entre la progenie (individuos 1 a 7) de una familia en la que no se conoce el genotipo materno. A, B y C representan bandas de distintos pesos moleculares



Estimaciones de la tasa de fecundación cruzada en poblaciones naturales de angiospermas

Un amplio porcentaje de las angiospermas son plantas hermafroditas que poseen diversos mecanismos florales que evaden los efectos negativos de la

endogamia. Por lo tanto, promueven la fecundación cruzada. Sin embargo, a pesar de las ventajas de la fecundación cruzada, la transición evolutiva más frecuente entre las estrategias reproductivas en las plantas es el cambio a sistemas predominantemente de autofertilización. En la literatura científica se ha propuesto que tanto la autofecundación como la fecundación cruzada pueden ser puntos estables y alternativos en la evolución de los sistemas reproductivos (Shemske & Lande, 1985; Barret y Eckert, 1990; ver Apéndice 1) y que muchos de ellos se encuentran en transición evolutiva hacia estos puntos alternativos (Charlesworth, 1982). Sin embargo, varias combinaciones espacio-temporales de las estructuras reproductivas (como el monoicismo), permiten que una proporción de la progenie sea producto de fecundación cruzada y otra de autofecundación. Por lo tanto, permiten las ventajas que se han propuesto para ambas estrategias reproductivas (fecundación cruzada y autofertilización). Por una parte, la progenie de fecundación cruzada tiene las ventajas de la recombinación (mayor variación genética), mientras que por otra, la progenie de autofertilización asegura el mantenimiento del genotipo parental, el cual generalmente se encuentra bien adaptado a las condiciones ecológicas prevalecientes. De esta forma, teóricamente una estrategia mixta puede ser favorecida y estable (Uyenoyama, 2000). De manera que en especies con sistemas mixtos, esperaríamos que la tasa de fecundación cruzada fuera intermedia. Hay que recordar que una estrategia mixta, al menos teóricamente, podría presentar ventajas selectivas al permitir más combinaciones genotípicas que podrían propagarse en diversas circunstancias ecológicas (Holsinger, 2000).

En los sistemas reproductivos asociados a un sistema de incompatibilidad, es esperable que la tasa de fecundación cruzada sea de uno o cercana a uno, mientras que en las especies autocompatibles (que aparentemente no poseen algún mecanismo para reducir la autofecundación), cercana a cero. Sin embargo, si los atributos del sistema reproductivo no se encuentran asociados a un sistema de incompatibilidad, la tasa de fecundación cruzada debería ser mixta. Si es así, la distribución de la tasa de fecundación cruzada resultaría en un continuo desde el 0 al 1, lo cual se contrapone a la distribución bimodal, reportada en otros trabajos (Barrett y Eckert, 1990; Barrett et al., 1996).

Con el fin de mostrar la relación de los sistemas reproductivos en las angiospermas y la tasa de fecundación cruzada, se realizó una revisión de la literatura publicada a partir de diferentes bases de datos. El criterio para incluir un artículo en nuestro análisis implicó el reporte de la tasa multilocus (t_m) estimada a partir de cualquier marcador molecular, así como la descripción

del sistema reproductivo y la presencia/ausencia de sistema de incompatibilidad. Los artículos seleccionados abarcaron un periodo entre 1974 y el 2006. Primero se generó una distribución de la tasa de entrecruzamiento multilocus para todas las especies sin considerar el sistema reproductivo (Figura 3). Posteriormente, se agruparon a las especies o poblaciones de una especie, en los diferentes sistemas reproductivos (Figura 4). Con base en la distribución realizamos un análisis de varianza de una vía (ANOVA) para evaluar si las diferencias en la tasa de fecundación cruzada promedio entre los sistemas reproductivos eran estadísticamente significativas.

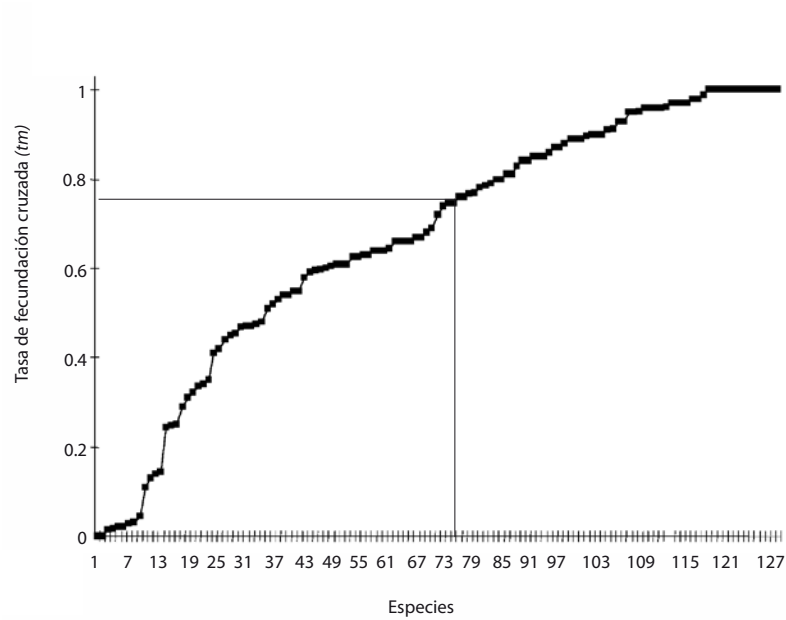
Se incluyeron un total de 131 trabajos que abarcaron 33 familias de angiospermas, 86 de los cuales fueron realizados en especies hermafroditas (47 autocompatibles y 39 autoincompatibles), 2 en especies androdioicas, 20 en ginodioicas, 7 en plantas enantiostilicas, 10 en heterostilicas (4 con especies distilicas, 3 con tristilicas y 3 con poblaciones monomorficas), 4 en especies monoicas y 2 con polimorfismo estilar. Las estimaciones fueron divididas en 20 categorías o sistemas reproductivos, algunas de las cuales corresponden a especies y otras a poblaciones de una especie con variación en el sistema reproductivo. Las enzimas fueron el marcador predominante (116 artículos), pues solamente 8 trabajos se realizaron con microsatélites, 2 con RAPDs y 2 con AFLPs.

El resultado más importante del análisis, fue el haber encontrado una distribución continua de t_m (Fig. 3) para 118 especies de angiospermas. Esta distribución contradice la hipótesis clásica que asume que la autofecundación y la fecundación cruzada son puntos alternativos en la evolución de los sistemas reproductivos en las angiospermas. Como puede verse en la Fig. 3, t_m no sólo sigue una distribución continua, sino que los valores intermedios son abundantes. Consideramos que el hecho de que tantas especies se encuentren en categorías intermedias significa que una estrategia mixta puede ser evolutivamente estable.

Aunque la tasa de fecundación cruzada promedio varió significativamente entre los sistemas reproductivos que analizamos, claramente no siguió una distribución bimodal, pues todos los sistemas reproductivos que no presentan un sistema de autoincompatibilidad, se encuentran en categorías intermedias de la distribución (Fig. 4).

El valor promedio más alto fue para las especies hermafroditas autoincompatibles, cercano a 1. El valor más pequeño es el de las especies monomórficas compatibles derivadas de especies heterostílicas. El resto de los sistemas que no difiere mucho del valor más pequeño podrían considerarse mixtos, ya que

Figura 3. Tasa de fecundación cruzada multilocus (t_m) de todas las especies sin importar el sistema reproductivo ni el sistema de incompatibilidad

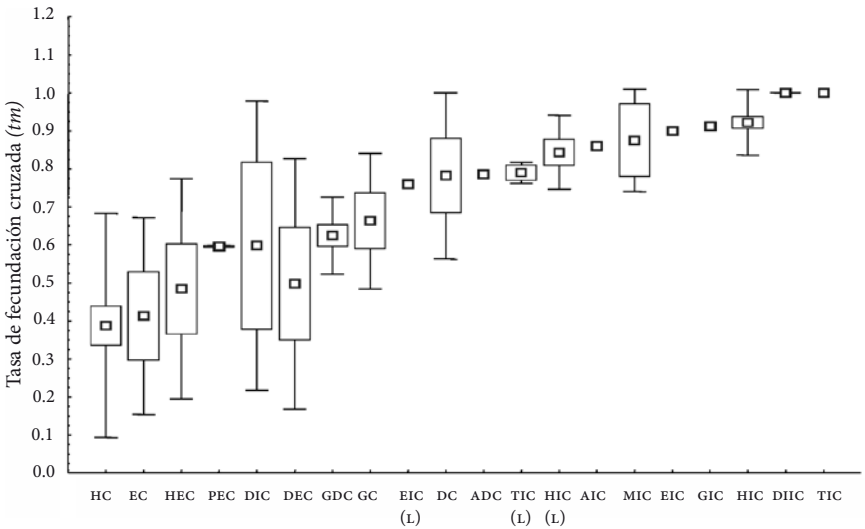


poseen valores entre 0.40 y 0.80 (Fig. 3). Algo que vale la pena resaltar es que las diferencias en la tasa de entrecruzamiento resultaron ser significativamente diferentes de cero (F = 4,2448; 6 g.l; $p < 0.0014$). Es decir, las características morfológicas y fisiológicas asociadas al sistema reproductivo repercuten directamente en la tasa de fecundación cruzada.

Debido a que en algunos de los sistemas reproductivos la muestra fue muy pequeña (por ejemplo PEC, ADC, DIIC y TIC) no fue posible realizar las pruebas de comparaciones múltiples para detectar las diferencias entre las categorías reproductivas. Por otra parte, en muy pocos trabajos se reportó la tasa de entrecruzamiento unilocus y la multilocus, cuya diferencia permite tener una estimación de los apareamientos entre parientes. Sin embargo, en las 6 especies en las que sí se reportó, varió de 0.06 a 0.38.

También resultó sorprendente que las especies hermafroditas autocompatibles presentaran una tasa de fecundación cruzada alta, resultado que podría ser interpretado de dos formas. Por una parte, el sistema reproductivo no ha

Figura 4. Tasa de fecundación cruzada multilocus (t_m) de veinte categorías en las cuales se agruparon distintos sistemas reproductivos tanto compatibles como incompatibles. HC= hermafrodita compatible. EC= enantiostílica compatbile. HEC= hermafrodita con hercogamia y compatible. PEC= hermafrodita con polimorfismo estilar y compatible. DIC= distílica y compatible. DEC= dicógama y compatible. GDC= ginodioica, dicógama y compatible. GC= ginodioica y compatible. EIC(L)= enantiostílica con incompatibilidad lábil. DC= dicógama y compatible. ACD= endrodióica, dicógama y compatible. TIC(L)= trstílica con incompatibilidad labil. AIC= androdioica incompatible. DIIC: distílica con incompatibilidad. TIC= tristilica con incompatibilidad. El cuadro representa la media, el rectángulo el error estándar y los bigotes la desviación estándar



sido descrito en detalle, por lo que es probable que en esta categoría estén mezclados varios mecanismos que promueven la fecundación cruzada. Por otra parte, Husband y Schemske (1996) revisaron la magnitud de la depresión por endogamia (δ) en plantas y encontraron que el promedio es de 0.23 en las especies que se autofecundan. Este resultado sugiere que las mutaciones con efectos deletéreos, aún en especies muy endogámicas, son difíciles de purgar. Por lo tanto, el muestreo de hijos en una familia podría estar sesgado a la progenie que proviene de fecundación cruzada, ya que es la que en promedio sobrevive más y esto sobreestimaría la tasa de fecundación cruzada.

Es importante mencionar que aunque existen varias revisiones sobre el tema, algunas de ellas congruentes con los resultados obtenidos aquí (Vogler y Kalisz, 2001), éste es el primer trabajo donde se analiza cómo varía la tasa de fecundación cruzada entre diferentes sistemas reproductivos. Una de las principales diferencias de este trabajo respecto a lo reportado en la literatura, está relacionado con la propuesta de Schemske y Lande (1985), que señala que el sistema de apareamiento en gran medida está determinado por la magnitud de la depresión por endogamia (δ). Teóricamente cuando $\delta < 0.5$ se espera que se favorezca la evolución de la autofecundación y cuando es mayor a 0.5, la fecundación cruzada. Es decir, en las especies hermafroditas autocompatibles una elevada tasa de entrecruzamiento podría estar determinada en parte por la depresión endogámica. Ellos pusieron a prueba su hipótesis y sugirieron que las categorías intermedias de *t*, pocas en su revisión, debían ser inestables y que se encontraban en transición evolutiva a un sistema reproductivo que promoviera la tasa de fecundación cruzada o a la autofecundación. Posteriormente los trabajos de Barrett y Eckert (1990) y Barrett *et al.* (1996) apoyaron esta hipótesis de la estabilidad de la autofecundación y la fecundación cruzada. Cabe hacer notar que en esos trabajos la tasa de fecundación cruzada se analizó de acuerdo al síndrome de polinización de las especies (concepto discutido en Herrera, 1996), agrupando en cada categoría muchos sistemas reproductivos. Esta manera de agrupar a las especies puede sesgar los datos hacia uno u otro lado de la distribución, debido a que la agrupación en síndromes de polinización no toma en cuenta atributos asociados a los sistemas reproductivos. En nuestra revisión agrupamos a las especies en sistemas reproductivos y estrategias como la hercogamia, dicogámia y el sistema de incompatibilidad que, como ya se mencionó, son mecanismos que disminuyen la endogamia y/o promueven la fertilización cruzada. Sugerimos que la interpretación de la tasa de fertilización cruzada debe tomar en cuenta estrategias reproductivas y no los síndromes de polinización.

Una tasa de fecundación cruzada mixta como la encontrada en este trabajo, teóricamente podría ser más estable que puntos alternativos como la autofecundación estricta o la diecia. Es decir, no existe una razón teórica clara para asumir que una estrategia reproductiva mixta es inestable. Ante los retos que impone la heterogeneidad ambiental, una población con una estrategia de reproducción mixta puede mantener las ventajas de una mayor diversidad genética producto del entrecruzamiento o bien asegurar su permanencia en la comunidad vía reproducción asexual. Por ejemplo, para *Phacelia dubia* y *Thymus vulgaris* (ginodioicas), se ha reportado variación en la magnitud de

(δ) asociada a las familias maternas. Sobre esta variación, la selección natural podría favorecer a ciertos genotipos en ciertas circunstancias ecológicas, y a otros cuando éstas cambiaran, generando un escenario de selección dependiente de las frecuencias. Ante escasez en el número de polinizadores, cuellos de botella o efectos de fundador serían favorecidos los genotipos en donde la (δ) sea menos severa (Del Castillo, 1998; Thompson y Tarayre, 2000).

En las angiospermas la endogamia biparental es el resultado de características florales y ecológicas. Por ejemplo, una dispersión limitada de polen y semillas, así como la interacción entre las plantas y sus polinizadores, influyen en la distribución espacial de los genotipos en la población y en el tipo de cruzas (Hardner *et al.*, 1998). La endogamia biparental es un componente que tiene un fuerte impacto en la dinámica evolutiva de las poblaciones al grado que contribuye a la depresión endogámica disminuyendo la viabilidad, el vigor o la adecuación de la progenie de cruzas entre individuos enparentados con respecto a la progenie de individuos no emparentados (Lewis, 1981; Fenester, 1991; Waser, 1993). Los efectos de la endogamia tienden a ser más severos en especies de fecundación cruzada obligada, en donde la endogamia biparental es el único modo de (δ), mientras que en especies autocompatibles o con sistemas mixtos esperariamos que fuera menor (Heywood, 1993). Debido a que existe una relación negativa entre (δ) y la adecuación de los individuos, la endogamia biparental podría tener un costo reproductivo alto. A pesar de la importancia de conocer la magnitud de t_m-t_s, ésta es reportada en mucho menor medida que la uniparental y las estimaciones indirectas no nos permiten conocer el grado de parentesco y la magnitud de la depresión por endogamia biparental y su relación con la evolución de los sistemas reproductivos en plantas.

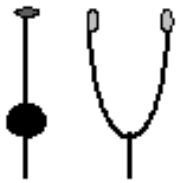
Encontramos algunos reportes de la tasa de entrecruzamiento con marcadores moleculares dominantes. Tradicionalmente, la tasa de entrecruzamiento se ha estimado a través de marcadores moleculares codominantes, particularmente enzimas, debido a que se pueden distinguir los tres genotipos (homócigos de ambos tipos y heterócigos). Sin embargo, existen especies en las que los niveles de variación genética con enzimas prácticamente no existen y la única forma de poder obtener la variación genética necesaria para una estimación como la tasa de fecundación cruzada, es a través de marcadores muy variables como los RAPDs, AFLPs o ISSRs. Sin embargo, hasta la fecha aún son pocos los estudios en donde se reporta dicha estimación, por lo que es muy difícil llegar a conclusiones generales.

Aunque la relación entre el sistema reproductivo y la tasa de fecundación cruzada están íntimamente ligados, no es muy común todavía encontrar es-

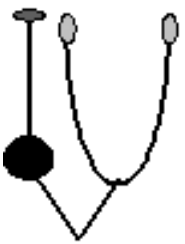
tudios o trabajos en donde se haya calculado la tasa de fecundación cruzada y, a su vez, caracterizado finamente el sistema reproductivo de una especie. Aún así, pensamos que la revisión aporta elementos para evaluar las repercusiones que podría tener la variación de un sistema reproductivo en la tasa de fecundación cruzada. Es decir, más que comportarse como entidades en transición a puntos estables y alternativos, tanto las especies con individuos que funcionan como entidades reproductivas discretas (como en la heterostilia), como aquellas que presentan variación cuantitativa, son más dinámicas espacial y temporalmente de lo que se ha asumido en los modelos teóricos e interpretado a partir de datos empíricos.

Apéndice 1. Definición de los sistemas reproductivos comúnmente descritos en la literatura y alguna de sus características generales

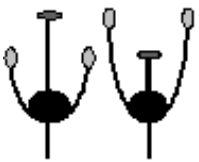
Diecia: población de plantas en donde una proporción de los individuos produce flores estaminadas o masculinas (generalmente el 50%) y la otra proporción produce flores pistiladas o femeninas. La diecia está considerada como la máxima separación de las funciones sexuales en plantas (Gebber, 1999).

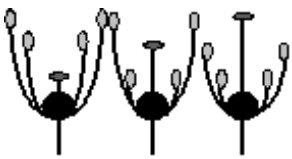


Monoecia: poblaciones de plantas en donde cada individuo posee flores pistilidas y estaminadas. A diferencia de la diecia, aquí las flores masculinas y femeninas están separadas dentro de un mismo individuo más que entre éstos. Algo interesante es que la diecia puede evolucionar a partir de la monoecia (Gebber, 1999).



Heterostilia: polimorfismo genético en el que dentro de una población coexisten dos (distilia) o tres (tristilia) morfos o formas florales que difieren recíprocamente en la longitud de las anteras y los estambres (hercogamia recíproca). Este polimorfismo además está acompañado por un sistema de autoincompatibilidad que evita la fecundación intra planta e intra morfo y una serie de características morfo-fisiológicas asociadas con las diferentes morfologías florales (Barret, 1992).

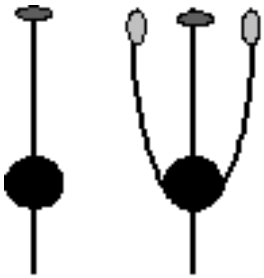




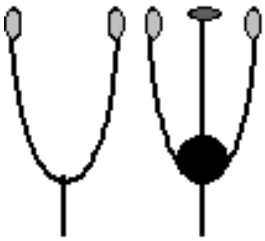
(Continúa)

Apéndice 1. Definición de los sistemas reproductivos comúnmente descritos en la literatura y alguna de sus características generales *(continúa)*

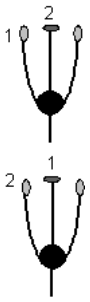
Ginodioicismo: Coexistencia de plantas hermafrodita y plantas hembra en una misma población. La frecuencia de hembras puede variar entre poblaciones y entre años dentro de una población. Además se ha propuesto que el ginodioicismo constituye uno de los primeros pasos y quizás el más importante en la evolución de la diecia (Gebber, 1999).



Androdioicismo: Coexistencia de plantas hermafrodita y plantas macho en una misma población. En un principio el androdioicismo también era considerado como una vía de evolución a la diecia. Sin embargo, ahora se sabe que la mayoría de las veces constituye una reversión de la diecia (Panel, 2002).



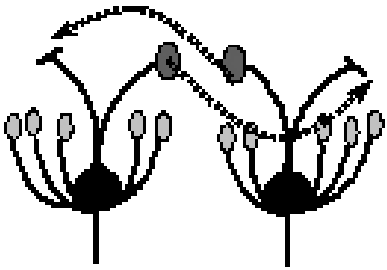
Dicogamia: término general utilizado para referirse a poblaciones de plantas hermafroditas donde las funciones sexuales masculina y femenina de cada individuo se separan temporalmente. Cuando la función sexual que primero madura es la masculina, se les denomina protándricas. Cuando por el contrario, la que primero madura es la femenina, entonces se les conoce como protogínicas (Gebber, 1999).



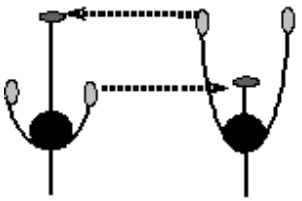
(Continúa)

Apéndice 1. Definición de los sistemas reproductivos comúnmente descritos en la literatura y alguna de sus características generales *(continúa)*

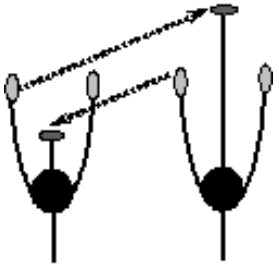
Enantostilia: Poblaciones de plantas compuestas de dos formas florales. Las formas florales están caracterizadas por presentar las estructuras reproductivas femeninas y masculinas de forma especular, como en un espejo. Hay especies en las que en cada planta se encuentran las dos formas florales en un mismo individuo, como en las monoicas y en otras por el contrario, se encuentran separadas en individuos, como en las dioicas. Prácticamente no existen trabajos al respecto (Barret, 2004).



Hercogamia: Separación espacial de las anteras y los estigmas en una misma flor. La hercogamia reciproca se refiere a la coincidencia en la misma posición del estigma y las anteras de dos tipos de flores ejemplo: las plantas heterostilicas.



Polimorfismo estilar: este polimorfismo hace referencia a poblaciones de plantas con diferencias en el largo de los estilos. Generalmente están compuestas de plantas con hercogamia aproximada donde los estigmas están por encima de las anteras y otro tipo de hercogamia menos frecuente es la reversa en donde los estigmas quedan por debajo de las anteras.



Sistema de incompatibilidad: Mecanismo fisiológico que impide la autofertilización. Esta basado en la habilidad de las células de la epidermis estigmática para distinguir la presencia de polen genéticamente similar, inhibir la germinación y el subsecuente desarrollo del tubo polínico (Nasrallah, 1997).

Bibliografía

Allard, R.W., y Workman, P. L. 1963. Population studies in predominantly self-pollinated species. IV. Seasonal fluctuations estimated values of genetic parameters in lima bean populations. *Evolution* 17: 470-480.

Awadalla, P., y Ritland, K. 1997. Microsatellite variation and evolution in the *Mimulus guttatus* species complex with contrasting mating systems. Molecular Biology and Evolution 14(10): 1023-1034.

Barrett, S.C.H., y Eckert, C.G. 1990. Variation and evolution of mating systems in seed plants. *In Biological approaches and evolutionary trends in plants*, edited by Kawano, London: Academic Press.

Barrett, S.C.H. 1992. Heterostylous genetic polymorphisms: Model systems for evolutionary analysis. En *evolution and function of heterosyly*, edited by Barrett, S.C.H. Springer-Verlag, Germany.

Barrett, S.C.H., Harder, L.D., y Worley A.C. 1996. The comparative biology of pollination and mating in flowering plants. Philosophical transactions in royal society of London, series B 351: 1271-1280.

Baudry, M., Kerdelhué, C., Innan, H., y Stephan, W. 2001. Species and recombination effects on DNA variability in the tomato genus. Genetics 158: 1725-1735.

Charlesworth, B. 1993. The evolution of sex and recombination in a warning environment. Journal of Heredity 84: 181-195.

Charlesworth, D., y Charlesworth, B. 1995. Quantitative genetics in plants: the effect of the breeding system on genetic variability. Evolution 49(5): 911-920.

Charleswoth, D., y Yang, Z. 1998. Allozyme diversity in *Lavenworthia* populations with different inbreeding levels. Genetics 81: 453-461.

Crow, J.F. 1988. *The importance of recombination in the evolution of sex: And examination of current ideas*, edited by Michod, R.E., and Levin, Sunderland: Sinauer.

Cuevas, E.G. 1996. Tasa de entrecruzamiento, vecindario genético y tamaño efectivo de la población en *Datura stramonium L.* (Solaneceae). Tesis de licenciatura, UNAM, México, D.F.

Darwin, C. 1877. *The different forms of flowers on plants of the same species*. John Murray: London.

Del Castillo R.F. 1998. Fitness consequences of maternal and nonmaternal components of inbreeding in the gynodioecious *Phacelia dubia.* Evolution 52(1): 44-60.

Dvorak, J., Luo, M., y Yang, L. 1998. Restriction fragment polymorphism and divergence in the genomic regions of high and low recombination in self-fertilizing and cross-fertilizing *Aegilops* species. Genetics 148: 423-434.

Fisher, R.A. 1930. *The genetical theory of natural selection*. Oxford: Clarendon.

Fenester C.B. 1991. Gene flow in *Chamaecrista fasiculata.* II. Gene establishment. Evolution 45: 410-422.

Geber, M.A. 1999. Theories of the evolution of sexual dimorphism. . En *Gender and sexual dimorphism in flowering plants*, edited by Geber A.M., Dawson, T.E., and Delph, L.F. Springer Verlag, Germany.

Hardner C.M., Potts, B.M., y Gore P.L. 1998. The relationship between cross success and spatial proximity of Eucalyptus globulus Ss.P. Globulus parents. Evolution 52(2): 614-618.

Hartl, D.L. 1980. *Principles of population genetics.* Sinauer associates; Sunderland, Massachusetts, U.S.A.

Hartl, D.L. 2000. *A primer of population genetics*, 3rd edition. Sinauer associates; Sunderland, Massachusetts, U.S.A.

Hayman, B.I. 1953. Mixed selfing and random mating when homozygotes are at a disadvantage. Heredity 7: 185-192.

Hedrick, P. 2000. *Genetics of populations*. Jones and Barlett Publisher: Boston.

Herrera, C.M. 1996. Floral traits and plant adaptation to insects: A devil's advocate approach. En *Floral biology: Studies on floral evolution in animal-pollinated plants*, edited by Lloyd, D.G., and Barrett, S.C.H. New York: Chapman and Hall.

Heywood, J.S. 1993. Biparental inbreeding depression in the self-incompatible annual plant *Gaillardia pulchella* (Asteraceae). American Journal of Botany 8(5): 545-550

Hickey, D.A. 2000. The evolution of sex and recombination. En *Evolutionary genetics: from molecules*, edited by Singh, R.S. and Krimbas, C.D. Cambridge University Press.

Holsinger, K.R. 2000. Reproductive systems and evolution in vascular plants. Proceedings of the National Academy of Science of the United States of North America, 97: 7037-7042.

Husband, B., y Schemske D. 1996. The evolution of the magnitude and timing of inbreeding depression in plants. Evolution 54: 54-70.

Kraft, T., Sall, T., Magnusson-Rading, I., Nilsson, N., y Halldén, C. 1998. Positive correlation between recombination rates and levels of genetic variation in natural populations of sea beet (*Beta vulgaris* subsp. *maritima*). Genetics 150: 1239-1244.

Levin, D.A. 1981. Dispersal versus gene flow in plants. Annals of Missouri Botanical Garden 68: 233-253

Lewontin, R.C. 1971. The effect of genetic linkage on the mean fitness of a population. Proceedings of the National Academy of Science of United States of America 68: 948-986.

Loyd, D.G., y Webb, C.J. 1992. The evolution of heterostyly. En *evolution and function of heterosyly*, edited by Barrett, S.C.H. Springer-Verlag, Germany.

Liu, F., Charlesworth, D., y Kreitman, M. 1999. The effect of mating system differences on the nucleotide diversity at the phosphoglucose isomerase in the plant genus *Leavenworthia*. Genetics 151: 343-357.

Maynard Smith, J. 1978. *The evolution of sex*. Cambridge: Cambridge University, Press.

Muller, H.J. 1932. Some genetic aspects of sex. American Naturalist 66: 118-138.

Nasrallah, J.B. 1997. Evolution of the Brassica self-incompatibility locus: a look into the S-locus gene polymorphism. Proceedings of the National Academy of Science of United States of America 94: 9516-9519

Pannell, J. 2002. The evolution and maintenance of androdioecy. Annual Review in Ecology and Systematics 33: 397-425.

Reynolds, J.D. 1996. Anima Breeding systems. Trends in Ecology and Evolution 2(2): 68-72.

Ritland, K., y Jain, S. 1981. A model for the estimation of outcrossing rate and gene frequencies using n independent loci. Heredity 47: 35-52

Ritland, K. 1983. Estimation of mating systems. En *Isozymes in plant genetics and breeding*, edited by Tanksley, S.D., and Orton, T.J. Elsevier, Amsterdam, Netherlands.

Roff, D.A. 2001. Inbreeding depression: test of the overdominance and partial dominance hypothesis. Evolution 56: 768-775.

Sakai, A.K., y Weller, G. 1999. Gender and sexual dimorphism in flowering plants: A review of terminology, biogeographic patterns, ecological correlates, and phylogenetic approaches. En *Gender and sexual dimorphism in flowering plants*, edited by Geber A.M., Dawson, T.E., and Delph, L.F. Springer Verlag, Germany.

Savolainen, O., Langley, C. H., Lazzaro, B. P., y Fréville, H. 2000. Contrasting patterns of nucleotide polymorphism at the alcohol dehydrogenase locus in the outcrossing *Arabidopsis lyrata* and the selfing *Arabidopsis thaliana.* Molecular Biology and Evolution 17(4): 645-655.

Schemske, D.W., y Lande, R.1985. The evolution of self-fertilization and inbreeding depression in plants. II. Empirical observations. Evolution 39: 41-52.

Schoen, D.J., y Brown A.1991. Intraspecific variation in population gene diversity and effective population size correlates with the mating system in plants. Proceedings of the National Academy of Science of United States of America 88: 4494-4497.

Thompson, J.D., y Tarayre, M. 2000. Exploring the genetic basis and proximate causes of female fertility advantage in gynodioecious *Thymus vulgaris*. Evolution 54: 1510-1520

Uyenoyama, M.K. 2000. The evolution of breeding system. En: *Evolutionary genetics from molecules to morphology.* Singh, R.S., Krimbas, C. (Eds.), Cambridge University Press. New York, pp: 300-316.

Vogler, D.W., and Kalisz, S. 2001. Sex among the flowers: the distribution of plant mating systems. Evolution 55(1): 202-204.

Waser, N.M. 1993. Sex, mating systems, inbreeding, and outbreeding. En *The natural history of inbreeding and outbreeding, theoretical and empirical perspectives*, edited by. Thornhill, N. W. TheUniversity of Chicago Press, Chicago.

Waser, N.M. 1993. Population structure, optimal outbreeding, and assortative mating in Angiosperms. En *The natural history of inbreeding and outbreeding, theoretical and empirical perspectives*, edited by. Thornhill, N. W. The University of Chicago Press, Chicago.

Webb, C.J. 1999. Empirical studies: Evolution and maintenance of dimorphic breeding systems. En *Gender and sexual dimorphism in flowering plants*, edited by Geber A.M., Dawson, T.E., and Delph, L.F. Springer Verlag, Germany.

Weir, B.S. 1996. *Genetic data analysis II*. Sinauer associates; Sunderland, Massachusetts, U.S.A.

Williams, G.C., Kubelik, A.R., Livak, K.J., Rafalski, J.A., y Tingey, S.U. 1990. DNA polymorphism amplified by arbitrary primers are useful as genetic markers. Nucleic acids research 18: 6531-6535.

Williams, G.C. 1975. *Sex and evolution*. Princeton, New Jersey: Princeton University, Press.

Wright, S. 1965. The interpretation of population structure by F-statistics with special regard to systems of mating. Evolution 19: 395-420.

Wright, S. I., Lauga, B., y Charlesworth, D. 2002. Rates and patterns of molecular evolution in inbred and outbred *Arabidopsis.* Molecular Biology and Evolution 19(9): 1407-1420.

Referencias incluidas en el análisis de la tasa de fertilización cruzada

Abbott R. y Gomez F. 1989. Population genetics structure and outcrossing rate of *Arabidopsis thaliana* (L.) Heynh. Heredity 62: 411-418.

Agren, J., y Schemske, D.W. 1993. Outcrossing rate and inbreeding depression in two annual monoecious herbs, *Begonia hisuta* and *B. semiovata*. Evolution. 47(1): 125-135.

Aide, M.T. 1986. The influence of wind and animal pollination variation in Outcrossing rates. Evolution 40 (2) pp 434-435.

Barrett, S.C.H., Cole, W.W., Arroyo, J., Cruzan, M.B. y Lloyd D. 1997. Sexual polimorphysms in *Narcissus triandrus* (Amaryllidaceae): is this species tristylous? Heredity 78: 135-145.

Barrett, S.C.H., y Shore J.S.1987. Variation and Evolution of Breeding System in the Turnera ulmifolia L. Complex (Turneraceae). Evolution 41(2): 340-354.

Belaoussoff, S., y Shore, J. S. 1995. Floral correlates and fitness consequences of mating-system variation in *Turnera ulmifolia*. Evolution 49(3): 545-556.

Brown, A.H.D., Zohary, D. y Nevo, E.1978. Outcrossing rates and heterozygosity in natural populations of *Hordeum spontaneum* Koch in Israel. Heredity 41(1) 49-62.

Brunet, J., y Eckert, C.G. 1998. Effect of floral morphology and display on outcrossing in Blue columbine, Aquilegia caerulea (Ranunculaceae). Functional ecology 12: 596-606.

Butcher, P.A. y Williams, E.R. 2002. Variation in outcrossing rates and growth in *Eucaplyptus camaldulensis* from the Petfrod region, Queensland; evidence of outbreeding depression. Silvae Genetica 51(1): 6-12

Cahalan, C.M., y Gliddon, C. 1985. Genetic neighborhood sizes in *Primula vulgaris.* Heredity 54: 65-70

Charlesworth, D., Morgan, M.T., y Charlesworth, B.1990. Inbreeding depression, genetic load, and the evolution of outcrossing rates in a multilocus system with no linkage. Evolution 44 6): 1469-1489.

Clegg, M.T., y Allard, R.W. 1972. Pattern of genetic differentiation in the Slender wild oat species *Avena barbata*. Proceedings of the Natural Academy of Sciences, USA 69: 1820-1824.

Culley, T.M., Weller, S.G., Sakai, A.K., y Rankin, A.E. 1999. Inbreeding depression and selfing rates in a self-compatible, hermaphroditic species, *Schiedea membranacea* (Caryophyllaceae). American Journal of Botany 86: 980-987.

Doligez, A. y Joly, H.I. 1997. Genetic diversity and spatial structure within a natural stand of a tropical forest tree species, *Carapa procera* (Meliaceae), in French Guiana. Heredity 79: 72-82

Ellstrand, N.C., Torres, A.M. y Levin, D.A.1978. Density and the rate of apparent outcrossing in *Helianthus annuus* (Asteraceae). Systematic Botany 3: 403-407.

England, P.R., Ayre D.J. y R.J. Whelan.1999. Microsatellites in the Australian shrub Grevillea macleayana (Protaceae). Primer notes. Molecular Ecology 8: 685-702.

Fenster, C.B. 1995. Mirror image flowers and their effect on outcrossing rate in *Chamaecrista fasiculata* (Leguminoceae). American Journal of Botany 82(1): 46-50.

Ferreira, J. J., Alvarez, E., Fueyo, M. A., Roca, A. y Giraldez, R. 2000. Determination of the outcrossing rate of *Phaseolus vulgaris* L. using seed protein markers. Euphytica 113: 259-263.

Fritsch, P. y Riesberg, L.H. 1992. High outcrossing rates maintain male and hermaphrodite individuals in populations of the flowering plant *Datisca glomerata*. Letters to Nature 359: 633-636

Gaiotto F.A., Bramucci, M., y Grattapaglia, D. 1997. Estimation of outcrossing rate in a breeding population of *Eucalyptus urophylla* with dominant RAPD and AFLP markers. Theoretical Applied Genetics 95: 842-849.

Glover, D.E., y Barrett, S.C.H. 1986. Variation in the mating of *Echinoria paniculata* (Spreng) Solms. (Pontederiaceae). Evolution 40: 1122-1131.

Gottlieb, L.D. 1975. Allelic diversity in the outcrossing annual plant *Stephanomeria exigua* SSP. Carotifera (Compositae). Evolution 29: 213-225

Hall, P.L.C., Orrell, y Bawa K.S. 1994. Genetic diversity and mating system in a tropical tree, *Carapa guianensis* (Meliaceae)

Hamrick, J.L. y Allard, R.W. 1972. Microgeographical Variation in Alloenzyme Frecuencies in *Avena barbata*. Proc. Nat. Acad. Sci. USA. 69 (8). 2100-2104.

Hayman, B.I. 1953. Mixed selfing and random mating when homozygotes are at a disadvantage. Heredity 7: 185-192.

Herlihy, C. y Eckert, C.G. 2004. Experimental dissection of inbreeding and its adaptative significance in a flowering plant, *Aquilegia Canadensis* (Ranunculaceae). Evolution 58(12): 2693-2703.

Hickey, D.A. 2000. The evolution of sex and recombination. In evolutionary genetics: from molecules, edited by Singh, R.S. and Krimbas, C.D. Cambridge University Press.

Holtsfrod T. y Ellstrand, N. 1989. Variation in Outcrossing rate and population genetic structure of *Clarkia tembloriensis*. (Onagraceae). Theorethical and Applyed Genetics 78:480-488.

Holtsfrod, T. y Ellstrand, N. 1992. Genetic and environmental variation in floral traits affecting outcrossing rate in *Clarkia temblorensis* (Onagraceae). Evolution 46: 216-225

Ishida, K. y Hiura, T. 2000. Mating system and population genetic structure of an androdioecious tree, *Fraxinus lanuginose* Koidz (Oleaceae) in northern Japan. Heredity 88: 296-301.

Iwaizumi M. G. y Sakai S. 2004. Variation in Flower biomass among nearby populations of impatiens textori (Balsaminaceae): effects of population plant densities. Canadian journal of botany. 82: 563-572

James T., Vege, S., Aldrich, P., y Hamrick, J. L. 1998. Mating systems of the three tropical dry forest three species. Biotropica 30(4): 587-594.

Jesson, L.K., y Barrett, S.C.H. 2002. The genetics of mirror-image flowers. Proceedings of the Royal Society of London, 269: 1835-1839.

Jesson, L.K. y Barrett, S.C.H. 2002. Enantiostyly in *Wachendorfia* (Haemodoraceae): The influence of reproductive system on the maintenance of the polymorphism. American Journal of Botany 89(2) pp: 253-262.

Karron, J.D., Jackson, R.T., Thumser, N.N., y Schlicht, S.L. 1997. Outcrossing rates of individual *Mimulus ringens* genets are correlated with anther-stigma separation. Heredity 79:365-370.

Kennington, J. W. y James, S.H. 1997. The effect of small populations size on the mating system of rare clonal mallee, *Eucalyptus argutifolia* (Myrtaceae). Heredity 78. pp. 252-260.

Kohn, J. R y Barrett S.C.H. 1994. Pollen Discounting and the spread of a selfing variant in tristylous *Echhornia paniculata*: Evidence from experimental populations. Evolution Vol 48. No 5. pp: 1576-1594.

Lande R., Schemske, D.W. y Schultz, S.T. 1994. High inbreeding depression, selective interference among loci, and the threshold selfing rate for purging recessive lethal mutations. Evolution 48 (4) pp: 965-978

Lewontin, R.C. 1971. The effect of genetic linkage on the mean fitness of a population. Proceedings of the National Academy of Science of United States of America, 68: 948-986.

Lu Yingking. 2000. Efects of density on mixed mating system and reproduction in natural populations of *impatines capensis*. Int. Plant Science 161(4): 671-681.

Mahy, G., Vekemans X. y Jacquemart, A.L. 1999. Patterns of allozymic variation within *calluna vulgaris* populations at seed bank and adult stages. Heredity 82: 432-440.

McKone, M.J. 1987. Sex allocation and outcrossing rate: a test of theoretical predictions using Bromegrasses (*Bromus*). Evolution 41(3): 591-598.

Medrano M., Alonso, C., y Herrera, C.M. 2005. Mating system, sex ratio, and persistence of females in the gynodioecious shrub *Daphne laureola* L. (Thymelaeaceae). Heredity 94: 37-43.

Medrano, M.C., Herrera C.M., y Barrett, S.C.H. 2005. Hercogamy and mating patterns in the self –compatible Daffodil Narcissus longispathus. Annal of botany 95: 1105-1111.

Moran, G. F., Muona, O., y Bell J. C. 1989. Breeding systems and genetic diversity in *Acacia auriculiformis* y *A. crassicarpa*. Biotrópica 21: 250-256.

Murawski, D.A., y Bawa, K.S. 1994. Genetic structure and mating system of *Stemnoporus oblongifolius* (Dipterocarpaceae) in Sri Lanka. American Journal of Botany 81(2): 155-160.

Murawski, D.A., Dayanandan, B. y Bawa K.S. 1994. Outcrossing raes of two endemic Shorea species from Sri Lankan tropical rain forest. Biotropica. 26: 23-29.

Norman J. K., Weller, S.G., y Sakai, A. 1997. Pollination biology and outcrossing rates in hermaphroditic *Schiedea lydgatei* (Caryophyllaceae). American Journal of Botany 84(5): 641-648.

Ohara, M., y Shimamoto, Y. 2002. Importance of genetic characterization and conservation of plant genetic resources: The breeding system and genetic diversity of wild soybean (*Glycine soja*). Plant species biology. 15: 51-58.

O´Malley, D.M., y Bawa, K.S. 1987. Mating systems of a tropical rain forest tree species. American journal of Botany 74(1): 1143-1149.

O´Malley, D.M., Buckley, D.P., Prance, G. T. y Bawa K. S.1988. Genetics of Brazil nut (*Bertholletia excelsa* Humb. & Bonpl.: Lecythidaceae). Theoretical applied genetics 76: 929-932.

Pérez-Nasser N., Eguiarte L. y Piñero D. 1993. Mating system and genetic structure of the distylous tropical tree *Psycotria Faxlucens* (Rubiaceae). American journal of Botany 80(1): 45-52.

Polowick, P.L., Vandenberg, A., y Mahon, J. D. Field assesment of outcrossing from transgenic pea (*Pisum sativum L.)* plants. Transgenic Research 11: 515-519 .

Reush T.B.H. 2000. Pollination in the marine realm: microsatellites reveal high outcrossing rates and multiple paternity in eelgrass *Zostera marina* Heredity 85: 459-464.

Rick C.M., Fobes, J.F., y Holle, M. 1977. Genetic variation in *Lycopersicon pimpinellifolium*: Evidence of evolutionary change in mating system. Plant Syst. Evol. 127:139-170.

Rick C. M., Holle M. y R.W. Thorpe. 1978. Rates of cross-pollination in *Lycopersicon pimpinellifolium*: impact of genetic variation in floral characters. Plant Syst. Evol. 129:31-44.

Runions J. C. y M. A. Geber. 2000. Evolution of the self-pollinating flower in *Clarkia Xantiana* (Onagraceae). I. Size and development of floral organs. American journal of Botany. 87(10): 1439-1451.

Rüter, B. J. L. Hamrick, y B. W. Wood. 2000. Outcrossing rates and relatedness estimates in pecan (Carya illinoinensis) populations. *The journal of heredity* 91: 72-75.

Sakai A.K., Weller S.G., Chen Mei-Ling, Yean-Chou S. y Tasanont C. 1997. Evolution of gynodioecy and maintenance of Females: The role of inbreeding depression, outcrossing rates, and resource allocation in *Schiedea adamantis* (Caryophyllaceae). Evolution 51 (3): 724-736.

Schmidt-Adam, G., Young A.G. y Murray, B.G. 2000. Low outcrossing rates and shift in pollinators in New Zeland pohutukawaw (*Metrosideros excelsa*; Myrtaceae). American Journal of Botany 87(9): 1265-1271.

Schoen, D. J., y Brown A. 1991. Intraspecific variation in population gene diversity and effective population size correlates with the mating system in plants. Proceedings of the National Academy of Sciences, USA. 88, 4494-4497.

Schoen D.J. 1982. The breeding system of *Gilia achilleifolia*: Variation in floral Characteristics and outcrossing rate. Evolution 36(2): 352-360.

Shaw D.V., Kahler, A.L. y E.W. Allard. 1981. A multilocus estimator of mating system parameters in plant populations. Proceedings of the National Academy of Sciences, USA. 78 (2): 1298-1302.

Soon-Leong Lee, Kevin K.S., Leng-Guan Saw N.G., Adnan Norwati, Siti Salwana M.H., Lee C.T., y Norwati, M. 2002. Population genetics of *intsia palembanica* (Leguminoseae) and Genetic conservation of virgin Jungle Reserves in Peninsular Malaysia. American Journal of Botany 89(3): 447-459.

Sun, M., y Ganders, F. R. 1986. Female frequencies in Gynodioecius populations correlated with selfing rates in hermaphrodites. American Journal of Botany. 73: 1645-1648.

Sun, M., y F. R. Ganders. 1988. Mixed mating system in Hawaiian Bidens (Asteraceae). Evolution Vol 42 No 3. pp. 516-527.

Sun, M., Wong, K. C. y Lee, J.S.Y. 1998. Reproductive biology and population genetics structure of *Kandelia candel* (Rhizophoraceae), a viviparous mangrove species. American Journal of Botany 85(11): 1631-1637.

Sun, M. y Ritland K. 1998. Mating system of yellow starthistle (Centaurea Solsitialis), a successful colonizer in North America. Heredity 80: 225-232.

Takebayashi, N., Wolf, D.E y Delph, L.F. 2006. Effect of variation in herkogamy on outcrossing within a population of *Gilia achilleifolia*. Heredity 96: 159-165.

Vasek, F.C., y J. Harding. 1976. Outcrossing in natural population . V. Analysis of out-crossing, inbreeding, and selection in *Clarkia exilis* and *C. tembloriensis*. Evolution 30: 403-411.

Van Dijk, H., Wolff, K., y De Vries, A. 1988. Genetic variability in Plantado species in relation to their ecology. III Genetic structure of populations of *P. major*, *P. lanceolata*, y *P. coronopus*. Theoretical Applied Genetics. 75: 518-528.

Van Treuren, R., Bijlsma, N.J., Ouborg N.J. y W. Van Delden. 1993. The Effects of populations size and plant density on outcrossing rates in locally endangered *Salvia pratensis*. Evolution. Vol. 47. No 4. pp.1094-1104.

Van Treuren R., Bijlsma N. J., Ouborg N. J y M. M. Kwak. 1994. Relationships between plant density, outcrossing rates and seed set in natural and experimental populations of *Scabiosa columbaria*. Journal of evolutionary biology 7: 287-302.

Wain, R. P. 1982. Genetic differentiation in the Florida subspecies of Helianthus debilis (Asteraceae). American journal of botany. 69: 1573-1578.

Wang G., Yamasue, Y., Itoh K. y Kusanagi, T. 1998. Outcrossing rates as affected by pollinators and the heterozygote adventage of *Monochoria Korsakowii*. Aquatic Botany 62: 135-143

Capítulo 7

La clonalidad y sus efectos en la biología de poblaciones

María C. Mandujano Sánchez

La clonalidad y sus efectos en la biología de poblaciones pueden entenderse si partimos del hecho de que buena parte de la teoría en ecología, genética, fisiología, conducta y evolución se ha desarrollado usando los individuos como punto focal; por ejemplo, la ecología se define como la ciencia que estudia las interacciones que regulan los patrones de distribución y abundancia de los individuos (Krebs, 1985). El desconocimiento de la biología de poblaciones de especies clonales estriba precisamente en que las teorías ecológicas y evolutivas se han desarrollado para entender organismos unitarios (Harper, 1977), mientras que las especies clonales forman más de un individuo con la misma composición genética, lo que dificulta la definición de individuo. Por ello los estudios de las plantas clonales se han desarrollado recientemente.

En sus inicios los trabajos genéticos y demográficos utilizaron como sistemas de estudio especies que no tuvieran "complicaciones": plantas de ciclo de vida corto (generalmente anuales), sin propagación vegetativa (aclonales) y hermafroditas (Bierzichudek, 1982). Ya que de manera general los modelos poblacionales y los genéticos se han desarrollado para especies unitarias, primordialmente animales (incluido el hombre), la problemática que se mencionará en este capítulo sobre estudiar o modelar la clonalidad podría generalizarse a otros atributos que dificultan el entendimiento de la biología de las especies, como el dioicismo, el cambio de sexo mediado por condiciones ambientales, la presencia de generaciones superpuestas, la

estructura poblacional por tamaño y la longevidad prolongada. La tendencia a estudiar organismos de ciclo de vida simple se refleja en la proporción de estudios demográficos y genéticos de especies con esas características que se puede observar en el trabajo de revisión de los atributos de historia de vida de 66 especies de plantas elaborado por Silvertown *et al.* (1993). Estos autores compilaron y analizaron información de todos los trabajos demográficos publicados hasta 1993, observando que en esta muestra de estudios el 68% de las especies son herbáceas y el 44% presentan clonalidad. Entre las causas de este sesgo destacan la facilidad para completar los estudios demográficos en corto tiempo y el uso directo de los modelos teóricos (basados en el individuo) para estimar la mortalidad y la adecuación. En cuanto a los estudios genéticos, gran parte de los trabajos que se han realizado con plantas clonales incluyen un muestreo dirigido (ver Silva y Eguiarte, 2003; Clark-Tapia *et al.*, 2005) con el objetivo de evitar en lo posible colectar individuos que presenten el mismo genotipo (ramets de un genet, ver más adelante). En este sentido, existe un rezago considerable en cómo modelar la biología de las poblaciones clonales, ya que inicialmente se consideró la clonalidad como un "ruido ambiental" que complicaba considerablemente la modelación de la biología de las poblaciones, ya que lleva a sobreestimar el número de individuos (genets) y porque los patrones de distribución espacial que se generan por la clonalidad afectan los supuestos de la reproducción aleatoria por dos mecanismos: aumentando la endogamia o disminuyendo la adecuación por depresión endogámica.

La conceptualización de lo que es un individuo es una de las cuestiones centrales en las plantas clonales. La principal separación se ha realizado entre los individuos unitarios y los individuos clonales (tabla 1).

Tabla 1. Características de los organismos unitarios, aclonales y clonales.
*Excepción son los gemelos univitelinos

Atributo	Organismo unitario	Organismo aclonal	Organismo clonal
Unidad de construcción	Unitario	Modular	Modular
Crecimiento	Determinado	Determinado o indeterminado	
	Indeterminado		
Integración (conexión física entre partes)	Sí	Sí	Potencialmente independientes

(Continúa)

Tabla 1. Características de los organismos unitarios, aclonales y clonales. *Excepción son los gemelos univitelinos *(continúa)*

Atributo	Organismo unitario	Organismo aclonal	Organismo clonal
Movimiento	Móviles	Sésiles	Sésiles y desplaza miento local
Dispersión de propágalos o vástagos	Grandes distancias	Grandes distancias	Localizada y a grandes distancias
Ejemplo	Animales, todos en la clase aves y mammalia*	Gimnospermas, manglares	Algas, pastos marinos, nopales

Si esta separación no es clara, las conclusiones que se obtengan de cualquier estudio tendrán problemas importantes. Por ejemplo, existen al menos cuatro aspectos en los que se ha identificado claramente esta problemática: i) los organismos clonales tienen la capacidad de fragmentarse en piezas más pequeñas, lo que hace que las estimaciones demográficas sean muy difíciles tanto de medir como de analizar; ii) conforme los clones se desarrollan y crecen se puede especular que la diversidad genética dentro de estas poblaciones disminuye; iii) la forma en la que los organismos clonales ocupan el espacio -si colonizan el espacio como falange o guerrilla (Harper, 1985) y si están conectados o desconectados- dificulta explicar su integración funcional o su sobrevivencia; y iv) por su demografía fragmentada y la acumulación de mutaciones somáticas es difícil establecer claramente cuál es la unidad de selección en estos organismos, por ejemplo se ha estimado que la longevidad de clones de *Larrea tridentata* hasta de 10,000 años (Vasek, 1980) y se ha observado que un mismo genet del nopal rastrero puede presentar ramets de nueve tallas distintas (Mandujano *et al.*, 2001).

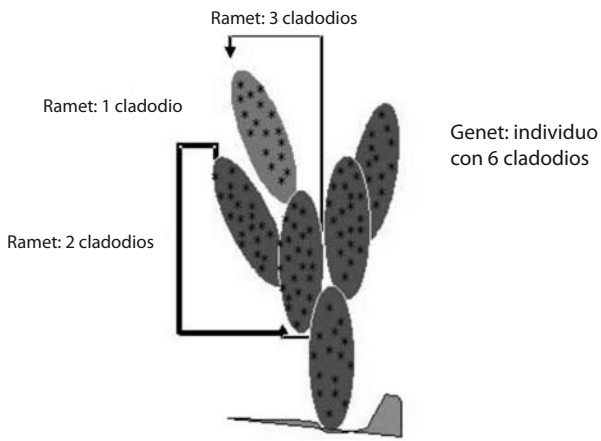
Organización jerárquica en plantas clonales y definiciones operativas

En la literatura existe mucha variación en la terminología de plantas clonales, por lo que resulta importante tratar de aclarar los conceptos. La confusión tiene un origen filosófico, ya que algunos autores consideran que la clonalidad es una forma de crecimiento, mientras que otros consideran que es una forma

de producir progenie. En esta sección se presentan diferentes definiciones de población, individuo, módulo, genet y ramet. Asimismo, dado que las especies clonales funcionan en diferentes etapas de su vida como conjuntos y que potencialmente pueden independizarse fisiológicamente, los organismos modulares, que son clonales, pueden estudiarse en distintos niveles de organización.

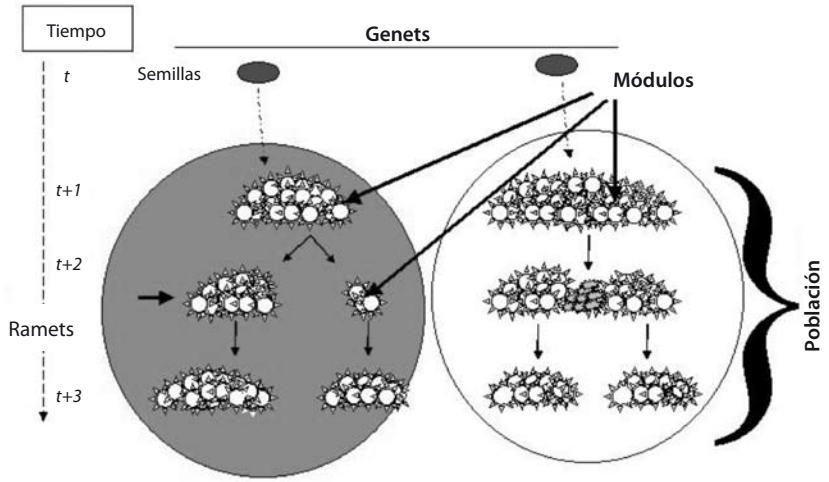
El *módulo* es la unidad de construcción básica que se repite o itera a medida que el individuo crece, es el nivel más sencillo; se definió originalmente como una hoja, una yema axilar y un entrenudo, pero dependiendo del organismo puede definirse y corresponder a otra estructura (Harper 1977, 1985). Así, el módulo apropiado para la descripción de un organismo como un conjunto de partes depende de la especie bajo estudio. Por ejemplo una penca o cladodio en *Opuntia* (Cactaceae), un tubérculo en *Mammillaria* (Cactaceae), un bulbo en una orquídea, entre otros (figura 1). El módulo *per se* o una agrupación de ellos puede originar una unidad, individuo u organismo potencialmente independiente (unidad denominada *ramet*). Todos los organismos que se formen (juntos o separados por cualquier mecanismo) producto de un cigoto forman un *genet* (algunos utilizan el término *clona*; Harper 1985; Jackson *et al.*, 1985; Tiffney y Niklas, 1985), que es un individuo con la misma composición genética (de cualquier tamaño y ocupando cualquier extensión), por lo que un ramet es una parte del genet.

Figura 1. Genet de un nopal hipotético (*Opuntia* spp.). El tamaño del genet es de 6 cladodios o pencas, cada cladodio es un ramet, aunque este nopal puede llegar a fragmentarse y formar ramets de diferente talla



El problema es en realidad definir unidades real o potencialmente independientes para hablar entonces de "crecimiento clonal" o reproducción o propagación vegetativa. Posiblemente la base para iniciar cualquier trabajo con especies clonales sea caracterizar con claridad el nivel en el que se elaborarán las conclusiones, ya sea poblaciones, genets, ramets o módulos (figura 2).

Figura 2. Niveles de organización en las plantas clonales. Se muestran los módulos, ramets, genets y una población modelo de 2 genets con una forma de vida cespitosa, similar a *Ferocactus robustus* (Cactaceae)



Dentro del círculo gris se ejemplifica una forma de clonación del tipo fisión o fragmentación a partir de una planta focal originada de semillas que se establece (tiempo *t*) y crece (tiempo *t*+1), posteriormente la planta establecida se divide en dos más pequeñas (tiempo *t*+2) y sigue su crecimiento a la siguiente estación (tiempo *t*+3). Todos los individuos dentro del circulo gris conforman el genet 1. En el círculo blanco se ejemplifica otra forma de fragmentación en la que la planta original nacida de semilla sufre muerte tisular de unos módulos (por envejecimiento, falta de recursos, enfermedad, etc.) y con el tiempo queda separada en dos individuos sin ninguna conexión. Todo los individuos que están en el círculo blanco son el genet 2. El número de individuos se modifica a través del tiempo, en este ejemplo no representamos nuevo reclutamiento vía semilla por lo que el número de genets es siempre igual a dos, el número de ramets se modifica en el tiempo, 3 ramets en *t* + 2, y cuatro ramets en *t* +3

Entenderemos a las *poblaciones* como un conjunto de individuos de la misma especie que comparten un espacio y un tiempo determinados y se reproducen entre sí (Begon *et al.*, 1996; Silvertwon y Lovett-Doust, 1993; Gotelli, 1995). Generalmente existe una definición práctica del individuo

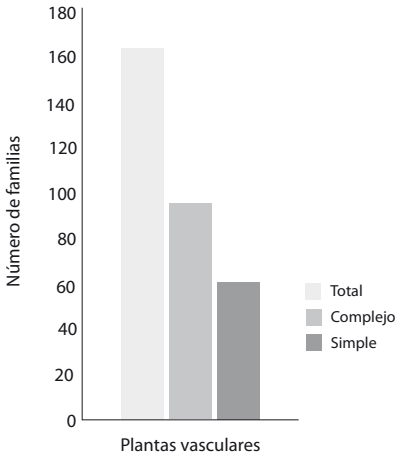
fisiológico que no está basada en un criterio genético estricto, en donde un individuo es cualquiera que se vea claramente separado como una unidad independiente en su parte vegetativa aérea (*sensu* Cook, 1979). De allí es evidente que en las poblaciones de especies clonales habrá más de un individuo fisiológico con un mismo genotipo, aunque posiblemente su talla tenderá a adjudicarle distintas posibilidades de sobrevivencia y fecundidad (Caswell, 1985). Otra propuesta interesante de individuo es la elaborada por Janzen (1977), quien sugiere que en las especies de organismos unitarios y en las aclonales existe un único individuo que es la unidad evolutiva; por el contrario, en las especies clonales existen varias unidades evolutivas, ya que considera implícitamente que habrá copias del individuo genético (ramets) que tendrán cada una un posible destino diferente como unidad evolutiva independiente.

La mayoría de los trabajos publicados sobre ecología de poblaciones de plantas clonales presentan análisis al nivel de ramet (Huenneke y Marks, 1987; Eriksson, 1988; Nault y Gagnon, 1993; Madujano *et al.*, 2001) ya que no se puede diferenciar un genet de un ramet si no se usan técnicas moleculares (ver Eriksson, 1993; Liston *et al.*, 2003; Clark Tapia *et al.*, 2005). Un número reducido de investigaciones se han realizado a nivel de demografía de módulos (*v.g.* López Portillo *et al.*, 1996; Golubov *et al.*, 2004) y la mayor parte de la investigación en plantas clonales es en el campo ecofisiológico (Watson, 1984; Wijesinghe, 1994; Falinska, 1995).

Tipos de clonación y frecuencia de clonalidad en las plantas

Tiffney y Niklas (1985) estimaron que el crecimiento clonal se presenta en aproximadamente 70% de las angiospermas (figura 3), más frecuentemente en las monocotiledóneas que en las dicotiledóneas, aunque se presenta también en muchas otras familias taxonómicas. Como resultado, las especies clonales son elementos importantes en todas las floras y están presentes en todos los hábitats, tanto terrestres como acuáticos. Sin embargo, la clonalidad puede presentarse como un requerimiento ontogenético en algunos taxa y el crecimiento clonal puede expresarse como una opción ecológica (Mogie y Hutchings, 1990). Es decir, aunque no sea parte de su programa de desarrollo, ciertas condiciones ambientales pueden favorecer o propiciar la aparición de la clonalidad en cualquier planta, debido a la presencia de tejido no diferenciado en los meristemos.

Figura 3. Más de la mitad de las familias de plantas vasculares cuentan con especies cuyos representantes tienen un ciclo de vida complejo, esto es, que su fecundidad y sobrevivencia dependen del tamaño o estado de desarrollo, la edad no afecta su conducta demográfica y pueden generar vástagos de origen sexual, asexual, y clonal



Independientemente del origen y la capacidad del taxón para clonar, los diferentes tipos de crecimiento se han agrupado en cuatro categorías: 1) crecimiento establecido, dado por el inicio del desarrollo de un organismo individual a partir de un propágulo; 2) crecimiento restaurativo, donde el daño local del individuo se repara -como sanar o cicatrizar una herida a través de la formación de tejido calloso; 3) crecimiento regenerativo, que involucra la reparación y el reemplazo de órganos dañados -como el crecimiento de una rama que se ha cortado; y 4) el crecimiento reproductivo, dado por la duplicación del individuo fisiológico (Tiffney y Niklas, 1985; tabla 2).

El crecimiento reproductivo o propagación vegetativa - que es la que nos interesa- puede ser de dos tipos: i) reiteración traumática, definida como la producción de un nuevo individuo inducida por estrés (como la producción de bulbilos en los magueyes por daño mecánico del escapo floral; Arizaga y Ezcurra, 2002) o bien ii) la reiteración programada, como la producción de un nuevo individuo por un meristemo como parte normal del desarrollo (en todas las especies con estolones, típico en gramíneas).

Estos tipos de producción de ramets pueden a su vez subdividirse en dos tipos de crecimiento con conexión y sin conexión a la planta madre

(*clonal growth linked and nonlinked*; Tiffney y Niklas, 1985). El crecimiento clonal sin conexión a la planta madre se caracteriza porque las estructuras pueden ser dispersadas más allá de la planta madre por vectores bióticos o abióticos, no por el crecimiento de una rama o raíz, y la descendencia no tiene una conexión vascular. La propagación sin conexión incluye todas las formas de pseudoviviparidad, bulbilos, frutos abortados o plantlets, la planta completa y semillas formadas por agamospermia, así como todas las estructuras que sean comparables con la semilla en cuanto a sus funciones ecológicas.

La propagación clonal sin conexiones se ha agrupado en cuatro grupos distintos: i) un meristemo axilar se desarrolla y forma a una planta adulta independiente de la madre; ii) gema/bulbilo, las nuevas plantas se producen a través del crecimiento de un callo, una unidad indiferenciada de tejido dispersable o del crecimiento de una planta pequeña especializada, diferente de la madre y capaz de ser dispersada - en ambos casos, la unidad de dispersión no es morfológicamente madura hasta que es dispersada y se establece; iii) la agamospermia, que incluye todas las formas de crecimiento clonal en las que no se presenta el proceso meiótico y se forman embriones diploides de manera asexual -en plantas con semillas un gametofito no reducido puede formarse directamente de tejido diploide del esporofito del óvulo o de una megaspora no reducida (Pimienta Barrios y del Castillo, 2002); y iv) la poliembrionía por división que se refiere a la producción de más de un embrión a partir de un solo cigoto fecundado (gemelos; Tiffney y Niklas, 1985; Pimienta Barrios y del Castillo, 2002). En algunas especies clonales las estructuras sexuales pueden modificarse y dar origen a un nuevo individuo vegetativo, un fenómeno denominado pseudoviviparidad. Este fenómeno se ha reportado para varias familias como Crassulaceae, Oxalidaceae, Polygonaceae, Saxifragaceae, Bromeliaceae, Poaceae, Juncaceae, Liliaceae y Gesneriaceae, y es muy común en ciertas especies de los géneros *Opuntia* y *Cylindropuntia* en Cactaceae (Anthony, 1954; Hicks y Mauchamp, 1999; Palleiro 2000; Palleiro *et al.*, 2006), en *Agave macroacantha* de la familia Agavaceae (Gentry, 1998; Arizaga y Ezcurra, 1995) así como en especies de ambientes extremos como el ártico y el desierto (Elmqvist y Cox, 1996). Las combinaciones de crecimiento reproductivo detalladas anteriormente nos arrojan un total de siete tipos de organismo clonal (tabla 2).

Tabla 2. Tipos de crecimiento reproductivo o propagación clonal.
*Reproducción asexual, en donde existe desarrollo embrionario
pero no hay recombinación, sólo es el genotipo materno

Crecimiento reproductivo		Tipo de organismo		
	i) Reiteración traumática	con conexión	Tocones de árboles	I
		sin conexión	Esquejes (rama o tallo que se introduce en la tierra y produce una nueva planta)	II
			Hojas (p. ej. violetas africanas)	
	ii) Reiteración programada	con conexión		III
		sin conexión	Meristemo axilar	IV
			Gema/bulbilo/plantlet	V
			Semilla* Agamospermia	VI
			Poliembrionía	VII

Ventajas y desventajas de la clonalidad

Las ventajas de la clonalidad son una ventana casi inexplorada hasta muy recientemente y producto de una visión más novedosa, en donde ciertos autores consideran que la clonalidad es una adaptación (Mogie y Hutchings 1990, Mandujano *et al.*, 2001). Esto se debe a que la clonalidad puede aparecer como una opción ecológica que permite alterar diversos atributos de historia de vida a favor de la dispersión de un genet (genotipo) en el tiempo y en el espacio. Por ejemplo, permite que algunas especies anuales, inmóviles y monocárpicas se transformen en especies perennes, "móviles" y policárpicas (Mogie y Hutchings, 1990). Un ejemplo es el de las especies del género *Agave*, cuya roseta es monocárpica pero al producir hijuelos vegetativos el genet es iteróparo (Arizaga y Ezcurra, 1995, 2002).

Otra ventaja es que es relativamente poco costosa, ya que no se necesitan los accesorios necesarios de la reproducción sexual para atraer a los polinizadores, como la producción de flores y recompensas como néctar, aromas

y polen para poder formar semillas; asimismo, permite mantener genotipos eficientes que han sido exitosos en el medio ambiente (ver Eguiarte *et al.*, 1999). En contraste, tiene la desventaja de que las plantas pueden tener una dispersión reducida (en comparación con las semillas), porque generalmente los propágulos clonales son más grandes y no cuentan con mecanismos desarrollados de dispersión. Sin embargo, esta limitación se puede librar produciendo semillas asexualmente como en la agamospermia o en algunos casos propágulos muy móviles como los de las especies acuáticas y marinas (Reusch *et al.*, 1998; Richards *et al.*, 2004).

Estudios detallados sobre la clonalidad en filogenias de muchos grupos sugieren que ha evolucionado más de una vez de manera independiente, que las estrategias de producción de tipo de propágulo son muy diversas y que no es un atributo ancestral, lo que apoya la idea de que la clonalidad es un carácter adaptativo (Tiffney y Niklas, 1985; Mogie y Hutchings, 1990).

Demografía: estructura y dinámica

Como se mencionó al inicio de este capítulo, el desconocimiento de la biología de poblaciones de especies clonales estriba en que las teorías ecológicas y evolutivas se han desarrollado para organismos unitarios. Como hemos visto, existen 3 niveles de organización en los que puede analizarse la demografía de una especie: el módulo, el ramet y el genet, aunque lo más frecuente es analizar la estructura poblacional en los dos últimos niveles. Una de las razones por las que la demografía de plantas no provee la misma información que la demografía de los animales, es porque en el primer caso, el conteo de individuos no es tan simple. En plantas no siempre estamos contando el número de individuos genéticos (genets) sino a los ramets, por lo que la tasa a la que la población crece o decrece siempre está sobreestimada. Éste es el gran problema en demografía de plantas clonales, por lo que entonces tenemos que definir el número de individuos o genets en nuestra población y el número de ramets que contiene. Sin embargo, la gran ventaja de estudiar la respuesta demográfica o de historia de vida a nivel de ramets es que podemos evaluar el desempeño del mismo genotipo en distintos escenarios: tamaños, edades y estados de desarrollo, entre otros.

Podemos hacernos exactamente la misma pregunta que en cualquier estudio de ecología de poblaciones: ¿qué determina el crecimiento de una población? sin olvidar que lo relevante es definir si nuestra población es de módulos, de ramets o de genets.

Los procesos que determinan el número de individuos en nuestra población serán la natalidad o número de individuos que nacen en una población (*B*), la mortalidad o número de individuos que mueren en la población (*D*) y la migración (individuos que inmigran *I* y los individuos que emigran *E*). Cabe mencionar que los procesos migratorios son generalmente ignorados en los modelos, ya que en la mayoría de los casos son muy difíciles de medir (véanse los capítulos 2 y 5 de este libro).

El modelo básico que determina el número de individuos en cualquier intervalo de tiempo (N_t) y que originalmente fue desarrollado para genets, pero que opera en la misma forma para ramets o módulos es (ecuación 1):

$N_{t+1} = N_t + B + I - D - E$ (ecuación 1)

$N_{t+1} = \lambda N_t$

Para cualquier tiempo: $N_t = \lambda^t N_{inicial}$ (ecuación 1.1)

$N = b - d - I - E$ (ecuación 2)

$N_t = N_0 e^{rt}$ (ecuación 2.1)

A partir del modelo sencillo expresado en la ecuación 1 se han desarrollado todos los modelos que permiten medir la tasa de crecimiento de cualquier población, generalmente haciendo la elección del modelo de acuerdo al ciclo de vida de la especie y a la duración del mismo, al momento en el que ocurre la reproducción, o si las generaciones son discretas o sobrepuestas. Por ejemplo, las expresiones más elementales para describir el número de individuos de cualquier población con crecimiento poblacional *discreto* son las ecuaciones 1 y 1.1 o con crecimiento poblacional *continuo* las ecuaciones 2 y 2.1. Las ecuaciones 1.1 y 2.1 se usan para la proyección del tamaño poblacional en el tiempo. Ambos modelos pueden posteriormente utilizarse de manera similar en poblaciones que se estructuran o se subdividen por edades, estadios de desarrollo o tamaño de los individuos (ver Caswell, 2001).

La tasa de crecimiento poblacional se puede calcular directamente a partir de estimar las tasas instantáneas de natalidad (*b*) y de mortalidad (*d*) de individuos $b=B/N$ y $d=D/N$, respectivamente. El parámetro maltusiano (*r* o tasa intrínseca de crecimiento poblacional) puede estimarse directamente para poblaciones no estructuradas como $r = (b-d)$ individuos/ind/tiempo. Por otro lado, la tasa finita de crecimiento poblacional denominada lambda (λ) puede estimarse al dividir el número de individuos en la población de un tiempo al

otro, donde $\lambda = N_{t+1}/N_t$ y fácilmente se puede ir de un modelo continuo a un modelo discreto con la relación entre las tasas de crecimiento de $\lambda = e^r$ ($\ln\lambda = r$) (Gotelli, 1995; Caswell, 2001). Ambos estimadores nos dicen si la población está creciendo. Si *r* vale cero, la población se mantiene con el mismo número de individuos (población en equilibrio); si *r* es mayor a cero, es decir positiva, su valor señala el número de individuos que se producen por cada individuo existente en la población en una unidad de tiempo; si *r* es negativa, entonces la población pierde los individuos que indica su valor, por cada individuo presente en la población, por eso es una tasa per cápita. Por su lado, lambda (λ) es una tasa adimensional (no tiene unidades), siempre es positiva y nos indica en qué proporción está cambiando la población, si es menor a uno la población disminuye y si es mayor a uno la población crece.

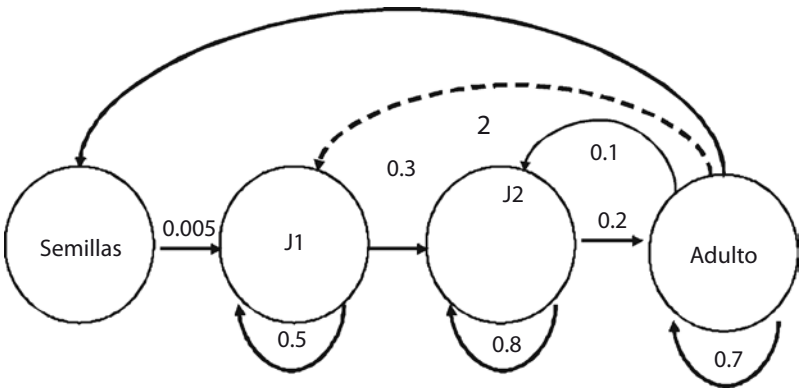
En poblaciones estructuradas se puede trabajar con tablas de vida (poblaciones estructuradas por edad; Gotelli 1995) o con modelos matriciales poblacionales (para cualquier tipo de estructura; por ejemplo diferente sexo o estadio de desarrollo; Caswell, 2001). Es importante señalar que en organismos clonales generalmente no se conoce la edad y que pueden permanecer en la misma categoría de estado o de tamaño de un tiempo a otro, por lo que el modelo adecuado es el de crecimiento discreto, matricial y de tipo Lefkovitch (1965; Mandujano *et al.*, 2001). Este modelo matricial permite que el crecimiento de la población se estime por la iteración de un vector de densidad población por una matriz de transiciones que resume los atributos demográficos (*power method*; Lefkovitch, 1965; Caswell 2001) como natalidad, fecundidad y sobrevivencia. Esta matriz se calcula con los datos del destino de individuos, etiquetados e identificables claramente, de una temporada de crecimiento a otra, generalmente anual, y sus fertilidades (figura 4, tabla 3; ver Golubov *et al.*, 1999, para el estudio de caso de un árbol; Mandujano *et al.*, 2001, de un nopal; Golubov *et al.*, 2004, de módulos). Este modelo sigue la ecuación:

$$n_{t+1} = A_{nt} \qquad \text{(ecuación 3)}$$

donde n_t son los vectores poblacionales en t y $t + 1$, y A es la matriz de transiciones demográficas.

Hay que tomar en cuenta que en los organismos clonales tenemos procesos demográficos diferentes a los que se presentan en organismos unitarios; estos procesos ocurren de una estación a otra (i.e., de t a t+1). Por ejemplo, la permanencia en el mismo estado de desarrollo (p. ej., plántula, juvenil, adulto); el crecimiento o salto más allá de la categoría de tamaño inmediata (p. ej. en

Figura 4. Este es el ciclo de vida de un cactus clonal estructurado en 4 estadios o categorias de desarrollo (círculos o nodos). La primera categoría corresponde a semillas, J1 y J2 son las categorías de juveniles y la cuarta categoría es de los adultos. Los adultos son los únicos que se reproducen; cada individuos produce 80 semillas (flecha sólida) y 2 hijos clonales de tamaño J1 (flecha punteada). Las flechas debajo de los nodos indican la posibilidad de que un individuo de la clase J1, J2 y adulto permanezca en el mismo estado de desarrollo de un tiempo *t* a otro *t+1*. Las flechas inmediatras entre un nodo y otro indican la probabilidad de tránsito o crecimiento de una categoría a la siguiente. Finalmente, la flecha de adulto a J2 indica el regreso de un adulto a la categoría J2 por achicamiento



épocas de muchos recursos los individuos tienen la capacidad de crecer aceleradamente, más que en una estación promedio); el encogimiento a cualquier talla menor (p. ej. si un individuo es comido por un herbívoro y pierde biomasa); el regreso a otro estado de desarrollo o la presencia de diferentes tipos de recién nacidos (p. ej. los individuos que se pueden reclutar en la población se originan de semilla, estolones, raíces, rosetas (chupones), cladodios caídos, frutos abortados o escapos dañados, entre otros; figura 4). El tipo de recluta, hijuelo, vástago o de recién nacido es importante demográficamente porque le puede conferir una adecuación diferencial, ya que se ha demostrado que los vástagos de origen sexual y clonal difieren en sus capacidades de sobrevivencia y reproducción, y presentan marcadas diferencias en cuanto al hábitat en que se establecen (Mandujano *et al.*, 1998). Las especies de nopal son ejemplos de organismos de ciclo de vida complejo: en *Opuntia rastrera* (Mandujano *et al.*, 2001) existen dificultades para caracterizar genets y ramets, los distintos tipos de recién nacido, las ventajas de los tipos de hijos en diferentes hábitats y además, existe la cuestión de la relación entre el tipo de hijo, el tamaño y el

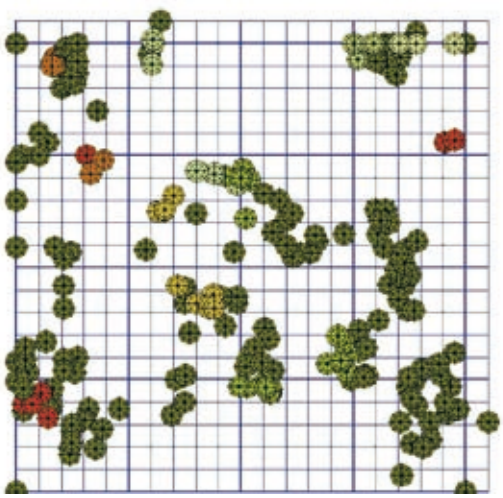
genotipo que se puede seleccionar (Caswell, 1985). El trabajo de Mandujano *et al.* (2001) es un ejemplo de cómo se integra en un modelo demográfico a la clonalidad (crecimiento vs. fecundidad clonal).

Distribución espacial de individuos en especies clonales y frecuencia de reclutamiento clonal en las poblaciones

El patrón de distribución espacial de los individuos en las poblaciones (uniforme, agregado o aleatorio) es uno de los temas más trabajados en ecología (MacArthur, 1984). Además de describir cómo es el patrón *per se*, se trata de explicar sus causas y sus consecuencias. Generalmente se ha encontrado que las especies de plantas clonales (primordialmente las terrestres) presentan un patrón agregado o contagioso (Reush *et al.*, 1998; Palleiro, 2002; Shimizu *et al.*, 2002; Carrillo Angeles, 2006) cuyas explicaciones están relacionadas con la baja dispersión de los propágulos (tabla 1). En este apartado es otra vez muy importante delimitar la categoría en la que vamos a describir el patrón, porque los ramets pueden presentar una distribución agregada, pero los genets podrían distribuirse en el espacio de una forma aleatoria (figura 5). También podría ocurrir que dos tipos de vástago presenten el mismo tipo de distribución espacial pero por causas ecológicamente distintas. Por ejemplo, los ramets de *Ferocactus robustus* se distribuyen de forma agregada porque los própagulos vegetativos se establecen cercanos a la planta progenitora, mientras que los genets también se distribuyen de manera agregada pero porque las plántulas se establecen asociadas a plantas nodrizas que las protegen de la radiación y de los herbívoros (Carrillo Angeles, 2006).

En general se espera que los vástagos que pueden dispersarse a grandes distancias (p. ej. semillas) den origen a patrones de distribución aleatorios. En este sentido, se sabe que el reclutamiento de individuos mediante reproducción sexual suele ser poco frecuente en numerosas especies clonales y puede llegar a no presentarse (Grant y Grant, 1980; Eriksson, 1988, 1993), particularmente en especies de suculentas que habitan en zonas áridas en donde 1 semilla de 14 millones que se produjeron sea la que llegue a establecerse (Mandujano *et al.*, 1998; Mandujano *et al.*, 2001). Muchas de estas plantas generan nuevos individuos principalmente a través de propagación vegetativa, la cual en una base anual es de 100 a 50% más frecuente que la reproducción sexual (Arizaga y Ezcurra, 1995; Mandujano, 1995; Mandujano *et al.*, 1996; Arizaga y Ezcurra, 2002) y generalmente produce patrones de distribución agregada.

Figura 5. El recuadro reticulado representa una parcela de 21 x 21 m (los cuadritos son de 1 x 1 m) donde se localizan los ramets y genets de *Opuntia rastrera* (Cactaceae). Los genets están representados por las sombrillas que comparten el mismo color. En este caso se observa que los mismos ramets están agrupados, mientras que el patrón de distribución de los genets no



Por otro lado, la clonalidad parece estar relacionada con ciertos ambientes caracterizados por presentar condiciones extremas y variables en el tiempo (Grime, 1982; Elmqvist y Cox, 1996). En estos ambientes la propagación vegetativa puede resultar ventajosa, ya que varios genotipos bien adaptados a las condiciones pueden mantenerse y los individuos originados por esta vía pueden tener menos dificultades para establecerse. Sin embargo, aunque la reproducción sexual puede provocar la pérdida de genotipos bien adaptados, es principalmente por este mecanismo que se forman nuevas combinaciones genéticas que pueden resultar a su vez altamente ventajosas en un ambiente particular (Willson, 1983; Eguiarte *et al.*, 1999; Kanno y Seiwa, 2004).

La distribución no aleatoria de los ramets en un área particular, como ya se ha dicho, provoca que los principales supuestos bajo los que operan tanto los modelos poblacionales como los genéticos se violen, en particular el de la panmixia. La expansión de clones en especies xenógamas (que se entrecruzan) representará una desventaja en comparación con la expansión de clones en las especies autógamas (que pueden autocruzarse), ya que la propagación clonal puede interferir con los patrones de dispersión de polen, afectando las

oportunidades de apareamiento de los genets (cruzas geitonogámicas intra-ramet e inter-ramet de un mismo genet). La geitonogamia en especies auto-incompatibles resultaría en la nula producción de frutos y semillas, mientras que en las especies autocompatibles llevaría a una reducción en la producción de frutos y semillas por depresión endogámica. El efecto de la distribución espacial de las plantas clonales sobre el sistema de cruza y apareamiento es uno de los campos de investigación novedosos para empezar a explorar (ver Handel, 1985; Charpentier, 2002) y sin duda nos permitirá entender cómo ha evolucionado la diversidad de historias de vida en los grupos de plantas con estas características.

Adecuación en plantas clonales

Podemos determinar la adecuación (*sensu lato*) de módulos, ramets o genets siguiendo la misma metodología. Puede intuirse que la más difícil de estimar es la adecuación de los genets, ya que requerimos forzosamente de la determinación de la identidad genética de los ramets para agruparlos en los distintos genets. Adicionalmente, la adecuación del genet será la adecuación de la sumatoria de la adecuación de los ramets que lo conforman, en un tiempo determinado. Esto está estrechamente ligado al patrón de distribución espacial que presenta el genet y los ramets, aunque solamente hay un trabajo con este enfoque (Carrillo Angeles, 2006).

De manera general, la adecuación puede calcularse a partir de la propuesta de la ecuación de Euler-Lotka, que es la más valiosa en el cálculo de la tasa per cápita de crecimiento poblacional y se obtiene a partir de la información de la tabla de vida. Sus principales supuestos son que las tasas de mortalidad y las fecundidades (específicas de cada edad) son constantes, y que la población ha alcanzado una distribución estable de edades (Hastings, 1996; ecuación 4):

$$\sum_{x=1}^{L} l_x \, m_x \, e^{-yx} = 1$$

donde l_x es la tasa de sobrevivencia desde el nacimiento hasta la clase de edad x, m_x es la fecundidad en esa clase de edad x, r es el parámetro maltusiano de crecimiento de la población, y e es el exponente o inverso de *ln*. Sin embargo, esta propuesta funciona para genets o bien para organismos unitarios, así que puede ser más operativo estimar la adecuación relativa de módulos, ramets o genets, que puede estimarse siguiendo la propuesta basada en Charlesworth (1980):

$$wH = \Sigma wH_i$$

donde *wH* representa la adecuación promedio del genet *H* y para conocerla debemos obtener la adecuación *w* de cada uno de los ramets (*i*-ésimo elemento con genotipo *H*) demográficamente equivalente a la tasa reproductiva neta denominada R_o, que puede ser estimada como $R_o = l_x m_x$, donde l_x es la sobrevivencia y m_x la fecundidad del ramet *x*. Esto puede reescalarse para módulos dentro de un ramet (figura 6).

Otra manera de estimar la adecuación es la propuesta de utilizar los coeficientes de sensibilidad como medida de la adecuación (van Groenendael *et al.*, 1988). La sensibilidad (s_{ij}) es la medida absoluta de un cambio producido en la tasa finita de crecimiento poblacional (λ) por un cambio en cada elemento (a_{ij}) de la matriz A (tabla 3). Caswell (2001) ha demostrado que los elementos de la matriz de sensibilidad (S=[s_{ij}]) pueden ser estimados directamente de los eigenvectores de A como: $s_{ij} = v_i w_j / w\, v$, donde v_i y w_j corresponden al *i*-ésimo y *j*-ésimo valor de los

Figura 6. El recuadro reticulado indica una parcela de 120 m^2, en donde se localizan los ramets de una especie ficticia. Los diferentes genets están representados por las sombrillas que comparten el mismo color y letra, el tamaño de cada sombrilla indica la talla del ramet. En este caso la adecuación $\bar{w}$ del genotipo *H* tendría que estimarse como la sumatoria de las adecuaciones parciales de cada ramet (ver texto).

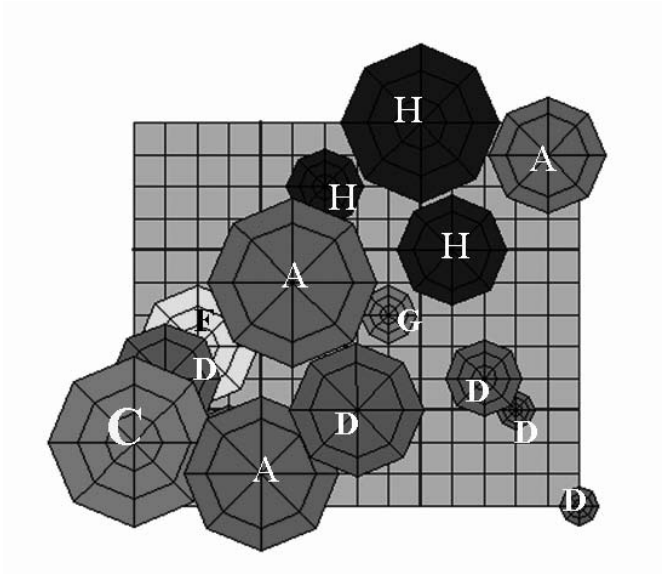


Tabla 3. Modelo de matriz de transiciones (**A**) que representa la contribución per capita de un individuo en cierto estadio o tamaño en el tiempo t a otro estado de los individuos en la siguiente generación $t + 1$ en una población

	Tiempo t (o contribución de la generación en t)							
	$\lambda = 1.21$	Semillas	Juveniles I	Juveniles II	Adultos	n_t	**w**	**v**
	Semillas	0			80	1000	0.91	1.00
Tiempo t+1	Juveniles I	0.005	0.5		2	505	0.04	0.05
	Juveniles II		0.3	0.8	0.1	150	0.04	0.04
	Adultos			0.2	0.7	500	0.01	0.02

Muestra las regiones y entradas (a_{ij}) de una matriz A de transiciones hipotética. El número de categorías (columnas j y renglones i, que siempre es simétrico) depende de la estructura poblacional de la especie de estudio. Las entradas o elementos a_{ij} indican a las variables demográficas como probabilidades o como número de vástagos y los subíndices dan la referencia a la columna y la fila a la que corresponden. Las regiones en la matriz representan los procesos demográficos y generalmente se conservan en las zonas que a continuación se señalan. En este ejemplo, los individuos de la población están agrupados en 4 categoría de estado (o en 4 estadíos). En el primer renglón las entradas o elementos de la matriz (a_{1j}) representan la fecundidad por vía sexual en número de vástagos de la categoría reproductiva de Adultos I a la categoría de semillas; los elementos diagonales ($a_{i=j}$) representan las probabilidades de quedarse del tiempo t a t+1 en el mismo estado de desarrollo, no crecer o mantenerse igual. Los elementos subdiagonales ($a_{i>j}$) representan el crecimiento y los elementos supradiagonales ($a_{i<j}$) pueden representar el encogimiento o retroceso a estados anteriores (p. ej. $a_{34} = 0.1$, indica que el 10% de los individuos adultos regresan a la categoría de Juveniles II de t a t+1) y la fecundidad por vía clonal (p. ej. $a_{24} = 2$, indica que los adultos producen en promedio 2 vástagos clonales de tipo Juveniles I de t a t+1). La columna n_t del cuadro representa el vector de densidades poblacionales, que indica el número de individuos presentes en cada categoría de estado en un tiempo t. El vector **w** (estructura estable en proporción) representa la proporción de individuos que se encuentra en cada categoría de estado en una población que ha alcanzado la estructura estable y **v** (valor reproductivo) representa el número de hijos que vale cada individuo en la categoría de estado i.

eigenvectores izquierdo y derecho respectivamente, y wv es el producto de estos dos eigenvectores que resulta en un escalar. La sensibilidad se obtiene de los elementos del vector de estructura poblacional (w) y de valor reproductivo (v) como:

$$s_{ij} = \alpha\lambda/\delta_{ij} = v_i w_j/\mathbf{vw}$$

La tasa de crecimiento poblacional (λ) y los vectores **w** y **v** se obtienen por diversos métodos, el más sencillo es el método de potencia (*power method*), que consiste en multiplicar la matriz **A** por el vector que contiene las densidades poblacionales $\boldsymbol{n_t}$ en un tiempo dado hasta que se alcance la estructura estable de edades (tabla 4). Numéricamente la estructura estable de edades (o de cualquier estructura de clase) se obtiene al dividir cada ***nj*** (el j-ésimo elemento de cada vector) de t+1 entre t hasta que la división de todos los elementos da el mismo valor. Es interesante que este valor al que "convergen" es λ (tablas 3 y 4). Otro método es elevar la matriz **A** a una potencia sufi-

Tabla 4. Estimación de lambda y del vector propio derecho (w, estructura estable de edades) utilizando el modelo de potencias (power method). La matriz de transiciones (A) que representa la contribución per capita de un individuo en cierto estadio o tamaño en el tiempo t a otro estado de los individuos en la siguiente generación t + 1 en una población. La columna nt del cuadro representa el vector de densidades poblacionales, el cual indica el número de individuos presentes en cada categoría de estado en un tiempo t. El vector w (estructura estable) representa el número de individuos que se encuentra en cada categoría de estado en una población que ha alcanzado la estructura estable (el mismo valor se representa en el Cuadro 3 como proporciones). El modelos se representa como n t +1 = A n t en donde A se multiplica por el vector nt, posteriormente A por el vector nt+1 y sucesivamente hasta que el modelo converge. La multiplicación de una matriz por un vector es la suma del producto del elemento renglón de la matriz por el elemento del vector columna y resulta en el elemento del siguiente vector [p. ej. el primer elemento del vector nt+1 ((0*1000) + (0*505) + (0*150) + (80*500)) = 40000]

Matriz de transiciones				Vectores de densidades poblacionales								
					n_t	n_{t+1}	n_{t+2} $_{nt+3}$	n_{t+4}	n_{t+5}	n_{t+6}	n_{t+7}	n_{t+8}
0	0	0	80	1000	40000	30400	26424	29256	37241	47776.8	59556.7	72707.4
0.005	0.5	0	2	505	1257.5	1588.8	1607	1667	1910.8	2336	2895.8	3563.4
0	0.3	0.8	0.1	150	321.5	672.5	1047.6	1356.8	1632.1	1938.6	2326.1	2820.5
0	0	0.2	0.7	500	380	330.3	365.7	465.5	597.2	744.5	908.8	1101.4
				nt+1/nt	nt+2/nt+1	nt+3/nt+2	nt+4/nt+3	nt+5/nt+4	nt+6/nt+5	nt+7/nt+6	nt+8/nt+7	
										Lambda*		
				40	0.8	0.9	1.1	1.3	1.3	1.2	1.2	
				2.5	1.3	1.0	1.0	1.1	1.2	1.2	1.2	
				2.1	2.1	1.6	1.3	1.2	1.2	1.2	1.2	
				0.8	0.9	1.1	1.3	1.3	1.2	1.2	1.2	

* Lambda equivale al valor del cociente n_{t+1}/n_t cuando converge al mismo valor para todas las densidades poblacionales (n_j)

cientemente grande (digamos 1000, $\mathbf{A}^{1000}$) pensando en que en ese número de iteraciones se alcanza la estructura estable y seguidamente se procede a multiplicar esa matriz resultante por un vector n_t. Los procedimientos referidos nos ayudan a estimar el vector de estructura estable (**w**). El vector **v** se estima siguiendo exactamente el mismo procedimiento pero ahora usando la matriz **A**' (**A**' transpuesta, que se obtiene cambiando los a_{ij} de renglones a columnas).Por definición la sensibilidad si contiene en una celda más de un proceso demográfico o más de un genet, $s_{ij} = (sh_{ij} + sk_{ij})$ (sensibilidad del genotipo *h* y *k*), la contribución de cada s_{ij} es la proporción del valor a_{ij} de la celda total (Flores-Martínez 1994; Mandujano *et al.*, 2001).

El problema es que este método solamente funciona si la especie no es clonal o si se ha construido una matriz de transiciones de cada genet. Lo anterior se debe a que los modelos matriciales están basados en el supuesto de que los individuos son unitarios. En el caso que nos ocupa, tendríamos que descomponer cada genet en una matriz **A**, como se muestra en el apartado de Análisis demográfico-genético, en donde se presenta una propuesta de cómo juntar la demografía y la genética y aplicar un método en donde se descompone la contribución proporcional de cada celda *aij* de una matriz **A** (tabla 3; Mandujano *et al.*, 2001).

Estructura y variación genética en especies de plantas clonales

En su concepto la genética de poblaciones en especies clonales no es distinta a la de los organismos unitarios, por lo que requerimos obtener estimadores de la diversidad genética como **Ht** (heterocigosis, como medida de diversidad genética total en una especie), **Hs** (diversidad genética dentro de poblaciones), **Gst** (diversidad genética entre poblaciones) y **Ps** (% loci polimórficos). Hamrick y Godt (1989) mostraron que las especies clonales presentan niveles altos de heterocigosis y que la diversidad genética no difiere entre especies que presentan reproducción sexual con respecto a las que presentan tanto sexual como clonal (tabla 5).

Sin embargo, estos índices de variación y diversidad genética suelen estimarse con métodos que evitan muestrear ramets (como selección de distancias *a priori*). Ellstrand & Roose (1987) propusieron un índice de diversidad genética particular para las plantas clonales, que se aboca a contestar cuál es la cantidad de genotipos que encontramos en una población y lo nombraron como la "proporción de descubrimiento de nuevos genotipos". Esta propor-

Tabla 5. Niveles de variación genética y su distribución entre poblaciones con diferente modo de reproducción, modificado de Hamrick y Godt (1989). *Ht* = diversidad genética total, *Hs* = diversiad genética dentro de poblaciones, *Gst* =diversidad genética entre poblaciones

Modo de reproducción	Taxa estudiados	Promedio de poblaciones	Promedio de loci	*Ps* (% loci polimórficos)	*Ht*	*Hs*	*Gst*
Sexual	352	13.1	16.7	51.6	0.311	0.229	0.225
Sexual y asexual	54	9.2	15	43.8	0.305	0.236	0.213

ción de genotipos distinguibles se denota *G/N,* donde *G*= número de clonas (genotipos distintos) y *N* = tamaño de muestra total (todos los individuos fisiológicos, ramets o genets) y toma valores de 0 a 1. Si *G/N* tiende a cero tenemos una población muy clonal, si por el contrario todos los genotipos son distintos (*G=N*), tenemos que *G/N* vale 1 y nuestra población u organismo no es clonal (tabla 6). Existen otros índices como el de diversidad de Simpson (Pielou, 1969):

$$D=1-\Sigma\,[n_i-(n_i-1))/(N(N-1))$$ (ecuación 7)

donde n_i es el número de ramets del *i*-ésimo genet y *N* es el número total de ramets. Asimismo se puede estimar la diversidad genética *D* en cualquier población, basada en el índice de Simpon con una corrección para poblaciones pequeñas $D=1-\Sigma\, p^2_i$ (ecuación 7.1; Molina Freaner y Clark Tapia, 2005).

De acuerdo con la hipótesis del beneficio de la reproducción sexual se espera que la diversidad clonal (*G/N*) sea mayor en las especies sexuales que en especies sexuales con dispersión restringida y menor que las que se reproduzcan principalmente de manera clonal. El intervalo reportado en especies clonales abarca de G/N= 0.01 (casi todo es un solo clon) a 1 (todos los individuos muestreados son genéticamente distintos), con un promedio de $G/N = 0.24 \pm 1$ e.e. 0.30. Sin embargo, hasta ahora la evidencia sólo permite concluir que las poblaciones clonales pueden ser tanto o más diversas que las sexuales (unitarias, tablas 2 y 3). Pero tenemos que considerar que existen sesgos en el muestreo, por lo que habrá que invertir en determinar los tamaños de muestra que nos puedan representar la escala de ramet y de genet de una manera significativa (por ejemplo Reusch *et al.*, 1998; Ivey y Rhichards, 2001; Shimizu *et al.*, 2002).

Tabla 6. Diversidad genética en especies de plantas clonales. MDVM=método de detección de la variación molecular, Ho= Heterocigosis observada, PE=número de poblaciones estudiadas. * Ellstrand & Rose, 1987 no indican los métodos usados en el trabajo original

Especies, familias y modo de vida	MDVM	PE	# de ramets	# de genets	G/N	Referencia
Anemone nemerosa (Ranunculaceae) Herbácea	Isoenzimas	20	600	568	0.95	Stehlik y Holderegger, 2000
Borrichia frutescens (Asteraceae) Herbácea rizomatosa	Isoenzimas	5	480	74	0.15	Richards *et al.,* 2004
Calystegia collina (Convolvulaceae) Herbácea	Isoenzimas	32	960	205	0.21	Wolf *et al.,* 1999
Carex bigelowii (Cyperaceae) Herbácea perenne	Isoenzimas	3	88	55	0.63	Jonsson *et al.* 1996
Carex curvula (Cyperaceae) Herbácea perenne	RAPD	1	116	15	0.13	Steinger *et al.* 1996
Cladium jamaicense (Cyperaceae) Pasto acuático	Isoenzimas	18	46	5	0.11	Ivey y Richards, 2001
Cryptomeria japonica (Cupressaceae) Leñosa	RAPD		263	64	0.24	Shimizu *et al.*, 2002
Cyperus esculentus (Cyperaceae) Herbácea	Isoenzimas		200	9	0.05	Horak *et al.*, 1987
Enteromorpha linza (Ulvaceae) Alga marina	*		1075	13	0.01	Ellstrand y Roose, 1987

(Continúa)

Tabla 6. Diversidad genética en especies de plantas clonales. MDVM=método de detección de la variación molecular, Ho= Heterocigosis observada, PE=número de poblaciones estudiadas. * Ellstrand & Rose, 1987 no indican los métodos usados en el trabajo original *(continúa)*

Especies, familias y modo de vida	MDVM	PE	# de ramets	# de genets	G/N	Referencia
Erigeron annus (Asteraceae) Herbácea	*		300	17	0.06	Ellstrand y Roose, 1987
Ferocactus robustus (Cactaceae) Globosa	RAPD	1	266	116	0.44	Carrillo y Mandujano, n.p.
Festuca rubra (Poaceae) Herbácea	*		1481	167	0.11	Ellstrand y Roose, 1987
Gaura biennis (Onagraceae) Arbustiva biannual	*		234	2	0.01	Ellstrand y Roose, 1987
Haloragodendron lucasii (Haloragaceae) Arbustiva	RAPD, isoenzimas	4	53	6	0.11	Sydes y Peakall, 1998
Haworthia reinwardtii (Liliaceae) Herbácea suculenta	*		145	18	0.12	Ellstrand y Roose, 1987
Lycopodium lucidulum (Lycopodiaceae) musgo	*		242	19	0.08	Ellstrand y Roose, 1987
Oenothera biennis (Onagraceae) Herbácea biannual	*		2200	46	0.02	Ellstrand y Roose, 1987
Poikilacanthus macranthus (Acanthaceae) Arbustiva	RAPD	4	277	68	0.25	Bush y Mulcahy, 1999

(Continúa)

Tabla 6. Diversidad genética en especies de plantas clonales. MDVM=método de detección de la variación molecular, Ho= Heterocigosis observada, PE=número de poblaciones estudiadas. * Ellstrand & Rose, 1987 no indican los métodos usados en el trabajo original *(continúa)*

Especies, familias y modo de vida	MDVM	PE	# de ramets	# de genets	G/N	Referencia
Populus tremuloides (Salicaceae) Leñosa	*		222	222	1.00	Ellstrand y Roose, 1987
Populus tremuloides (Salicaceae)	Micro-satelitales	4	15	11	0.73	Wyman *et al.*, 2003
Pueraria lobata (Fabaceae) Trepadora perenne	Isoenzimas	20	960	202	0.21	Pappert *et al.*, 2000
Salix exigua (Salicaceae) Leñosa	AFLP		49	11	0.22	Douhovinkoff y Dodd, 2003
Saxifraga cernua (Saxifragaceae) Herbácea	RAPD	2	93	13	0.14	Gabrielsen y Brochmann, 1998
Spartina alterniflora (Poaceae) Pasto marino	Isoenzimas	5	480	447	0.93	Richards *et al.,* 2004
Spartina patens (Poaceae) Pasto marino	Isoenzimas		346	101	0.29	Silander, 1984
Stenocereus eruca (Cactaceae) Columnar postrada	RAPD, isoenzimas	4	120	100	0.83	Molina Freaner y Clark-Tapia, 2005
Stenocereus gummosus (Cactaceae) Columnar postrada	Isoenzimas					Molina Freaner y Clark-Tapia, 2005

(Continúa)

Tabla 6. Diversidad genética en especies de plantas clonales. MDVM=método de detección de la variación molecular, Ho= Heterocigosis observada, PE=número de poblaciones estudiadas. * Ellstrand & Rose, 1987 no indican los métodos usados en el trabajo original *(continúa)*

Especies, familias y modo de vida	MDVM	PE	# de ramets	# de genets	G/N	Referencia
Taraxacum obliquum (Asteraceae) Herbácea	*	12	64	1	0.02	Ellstrand y Roose 1987
Trifolium repens (Fabaceae) Herbácea	*		1487	11	0.01	Ellstrand y Roose 1987
Vaccinium stamineum (Ericaceae) Arbustiva	RAPD		99	67	0.68	Kreher, Fore y Collins, 2000
Wyethia reticulata (Asteraceae) Herbácea	Isoenzimas, RAPD	2	29	8 0.28		Ayres y Ryan, 1997
Zostera marina (Zosteracae) Pasto marino	Micro-satélites	1	95	54	0.57	Reusch *et al.*, 1998
Promedio para todas las especies					0.29	

Resultado del resumen que se presenta en la tabla 6 podemos concluir que los marcadores que nos dan mayor información y de uso más generalizado en la determinación de genets son los RAPDs (*random amplified polymorphic DNA*; Hadrys, *et al.*, 1992; véase capítulo 18 de este libro). Los RAPDs son secuencias de DNA repetido o azaroso, con altas tasas de variación, que nos dan mucha información con un solo oligonucleótido y son relativamente baratos y rápidos. Otro método equivalente es el uso de marcadores hipervariables de secuencias repetidas (*intersequence simple repeats* ISSR; Wolfe *et al.*, 1998; tabla 5, véase capítulo 19 de este libro).Una ventaja de los ISSRs es que son más repetibles y las electroforesis pueden hacerse tanto en geles de agarosa y teñirse con bromuro de etidio como en geles de acrilamida con tinción de nitrato de plata. Finalmente, con cualquiera de los dos marcadores, los genotipos

de los ramets muestreados se evalúan comparando los patrones de bandeo (por presencia / ausencia de bandas) entre todos los individuos, utilizando las distancias genéticas de Nei-Li (1979) calculadas como $SG = 2m_{xy}/(m_x+m_y)$, donde m_{xy} es el número total de bandas compartidas entre pares de individuos, mientras m_x y m_y son el número total de bandas presentes en una muestra. Esta medida permite definir si cada par de individuos analizados tienen el mismo genotipo o corresponden a un mismo genet, lo que puede determinarse con base en tres criterios, que $SG = 1$ (i.e., todas las bandas son iguales entre los dos individuos), $SG = .95$ a 1 (entre 95 y 100% de las bandas son iguales) y $SG = .75$ a 1 (del 75 al 100% de las bandas son iguales). Este último criterio permite salvar el error de similitud o parentesco por endogamia o por mutación somática, aunque la elección del criterio a seguir también dependerá de las características de la especie bajo estudio.

Los datos publicados manejan en general rangos de 10 a 3 primers de 21 a 50 pares de bases, en muestras grandes (de 50 a ca. 300 individuos, referencias en la tabla 6). Esto muestra por un lado que en especies con polimorfismo es fácil caracterizar los ramets y que en las que la clonalidad no es muy frecuente (o quizá la endogamia), con muy pocos primers se puede obtener la identidad de genotipos y a partir de ese punto, calcular los índices de diversidad genética.

Análisis demográfico-genético

Finalmente voy a presentar una propuesta de cómo se puede vincular a la genética y la demografía. Si contamos con buenos datos demográficos y la identidad de los genotipos podremos juntar esa información en la construcción de modelos matriciales. Cabe señalar que sin duda es una perspectiva demográfica. Para cada población estudiada de nuestra especie clonal el análisis formal de los ciclos de vida anuales (o promedio de varios años de medición, de Steven 1989; Mandujano *et al.*, 2001) puede hacerse utilizando modelos poblacionales matriciales determinísticos, en los que el crecimiento de la población por unidad de tiempo es el producto de la matriz **A** por el vector $\mathbf{n}_t$ como se mostró en la ecuación (3) y en los tablas 3 y 4 (Caswell, 2001). En esta sección se define una matriz **B**, que es equivalente a la matriz **A**, pero la denotaremos con otra letra para indicar que representa la demografía de un genet diferente.

Se debe calcular una matriz de transiciones para cada genotipo ($\mathbf{B}_1, \mathbf{B}_2,..,\mathbf{B}_n$), con las mismas dimensiones o categorías de talla para que puedan sumarse, y

todas las matrices se sumarán para obtener una matriz de todos los genotipos para la población total ($\mathbf{A} = \Sigma (\mathbf{B_1}, \mathbf{B_2},..,\mathbf{B_n})$) (ecuación 8) por hábitat, por año o las condiciones que estemos evaluando en nuestra especie. La matriz **A** de genotipos se utilizará para estimar la tasa de incremento poblacional anual. Las entradas de las matrices de los genotipos se calcularán considerando el tamaño de la población por clase de tamaño.

La información de una matriz de transición **A** está organizada por contribuciones de la columna (*j*) al renglón (*i*), en donde el primer renglón corresponde a la fecundidad F_{1j} = número de semillas producidas por individuo en la clase de tamaño *j*. La diagonal principal contiene las probabilidades de sobrevivencia en la misma clase de tamaño (P_{ii}). Las subdiagonales contienen las probabilidades de crecimiento (G_{ij}), y las diagonales superiores corresponden a la fecundidad vegetativa (V_{ij} ; número de individuos vegetativos de tamaño *i* producidos por un individuo en la clase de tamaño *j*) y el regreso de individuos a categorías de menor tamaño (R_{ij} = número de individuos de tamaño *i* al tiempo $t + 1$, que estaban en la categoría *j* al tiempo *t* / número total de individuos en la clase *j* al tiempo *t*, donde $i<j$; Mandujano *et al.*, 2001, figura 4). Cabe hacer notar que todas las matrices tienen la misma forma que la expresada en la tabla 3, lo que cambia es la construcción de la misma.

Para cada tipo de vegetación o hábitat se construirán matrices anuales, basadas en los datos de cada temporada de crecimiento. El análisis matricial se puede realizar usando el método de potencias o con el programa MATLAB. Entre otros parámetros demográficos se puede estimar la tasa finita de crecimiento (λ, el eigenvalor dominante de la matriz), la estructura estable de edades (**w**, eigenvector derecho asociado a λ), y el valor reproductivo (**v**, vector izquierdo asociado a λ). Las contribuciones absolutas y relativas a λ de los procesos de crecimiento, permanencia, regreso a clases menores, reproducción sexual y propagación vegetativa se pueden evaluar por análisis de sensibilidad y elasticidad (Caswell, 2001; de Kroon *et al.*, 1986) y por análisis de rutas (van Groenendael *et al.*, 1994). Como ya se mencionó, la sensibilidad es una medida absoluta del cambio producido en lambda por un cambio en cada uno de los elementos a_{ij} de la matiz **A**, formalmente definido en la ecuación 6. Las sensibilidades pueden también interpretarse como una medida de los coeficientes de selección natural y en especial cuando se conoce la identidad genética de los individuos (van Groenendael *et al.*, 1988). Por otro lado, la elasticidad ($\mathbf{E}=[e_{ij}]$), mide la sensibilidad de λ a cambios proporcionales en los coeficientes de la matriz **A**, y se define como: $e_{ij} = (a_{ij}\ \partial\lambda)/(\lambda\ \partial a_{ij})$, y puede ser calculada a partir de los elementos de la matriz **S** como: $e_{ij} = (a_{ij}/\lambda)\ (s_{ij})$ (ecuación 9).

Las elasticidades suman uno (($\sum\sum e_{ij}$=1; Kroon *et al.*, 1986; Mesterton-Gibbons, 1993), lo que vuelve este análisis útil para estudios comparativos, como el que se describe aquí. Los elementos de la matriz de elasticidad pueden ser sumados por clase de tamaño o por proceso demográfico (*e.g.* Gotelli, 1991; Silvertown *et al.*, 1993). Cuando una celda contiene más de un proceso demográfico, como en el caso de los regresos y la propagación vegetativa, y cuando una celda contiene más de un genotipo, sus elasticidades son simplemente las proporciones correspondientes a los valores de la celda en la matriz de transición (Flores-Martínez, 1994; Mandujano *et al.*, 2001).

Las matrices de sensibilidad y elasticidad se deben calcular para todas las matrices anuales y se realizará un análisis de ruta de cada genotipo. El análisis de ruta es una extensión de los análisis de elasticidad (Van Groenendael *et al.*, 1994) y ha sido utilizado para explorar la ruta que más contribuye a la tasa de crecimiento poblacional (Nault y Gagnon, 1993). La extensión que se propone ofrece la oportunidad de comparar las rutas por las cuales los diferentes genotipos pueden completar su ciclo de vida. Por ejemplo, el aplicarlo sobre las matrices de genotipos en varios años (variación de condiciones ambientales) permite estimar el éxito relativo de los genotipos en términos de reproducción sexual, propagación vegetativa o la susceptibilidad de muerte en distintos años y ambientes.

Discusión y consideraciones finales

Para entender la clonalidad, la teoría ecológica y evolutiva de biología de poblaciones desarrollada para organismos unitarios, con poblaciones conformadas por individuos derivados de reproducción sexual, con crecimiento y longevidad determinados, debe dar una mirada a la importancia -comparativamente hablando- del número de especies con ciclo de vida complejo que buscamos comprender. Se debe por lo tanto reconocer que las plantas clonales han de estudiarse bajo el esquema de la definición práctica del individuo fisiológico, que no está basada en un criterio genético, en la que un individuo es cualquiera que se vea claramente separado como una unidad independiente en su parte vegetativa aérea (*sensu* Cook, 1985). Por ejemplo, si la especie es muy clonal podríamos encontrar baja diversidad genética. Sin embargo, el individuo fisiológico podrá enfrentar destinos distintos, en donde será relevante determinar cómo es la adecuación a la luz de la interacción genotipo por talla o estado de desarrollo.

En contraste al rezago empírico se ha establecido un desarrollo teórico extenso sobre cuáles son las ventajas y desventajas evolutivas de la presencia de la clonalidad, sobre todo para tratar de entender cómo se mantiene y cómo es que se ha convertido en una "estrategia" tan frecuente (Jackson *et al.*, 1985). Además, diversos autores comparan las ventajas y desventajas evolutivas de la clonalidad con respecto a la sexualidad, la longevidad de especies clonales y se ha trabajado muy extensamente en aspectos ecofisiológicos que incluyen cómo se integran y funcionan las plantas clonales (Williams, 1975; Willson 1979; Jackson *et al.*, 1985; Bell, 1987; de Kroon y van Groenenael, 1997).

Posiblemente la mayor limitante para entender cómo funcionan y evolucionan las plantas clonales ha sido la carencia de herramientas genéticas que permitan discernir si un individuo pertenece a un clon (ramet) o es un individuo genéticamente diferente. Esta simple separación que a nivel teórico es muy simple, genet o ramet, operativamente en las poblaciones naturales es, con algunas claras excepciones, imposible de evaluar si no se hace un estudio genético. Sin embargo, aún queda mucho camino por andar para estudiar los organismos clonales con una visión contrastante a los organismos unitarios. Esto está relacionado con las corrientes que conciben la reproducción y la propagación clonal como crecimiento o como una forma de producir propágulos; mi experiencia me hace inclinarme por la última.

Finalmente, pensemos en la paradoja de los gemelos de Einstein (Resnick y Halliday, 1992): de dos gemelos idénticos, uno es enviado en un viaje al espacio y vuelve, mientras que el otro permaneció en la Tierra. El gemelo que viaja lo hace a ¾ la velocidad de la luz y por el efecto de la dilatación de tiempo debida a la relatividad, el gemelo que partió experimenta una disminución del paso del tiempo y al volver será 4 años más joven que el gemelo que se quedó en la Tierra. El hecho que nos interesa es que aún cuando estos dos gemelos tienen un genotipo idéntico, la situación que viven determinará que uno de ellos sea más joven, mientras que el que se quedó en la Tierra será 4 años más viejo. Así tendremos historias distintas para cada ramet que seguramente conferirán al individuo fisiológico destinos contrastantes. Finalmente, para entender la clonalidad, pensaremos cómo será la historia de un cigoto que puede producir un número indeterminado de propágulos y dispersarlos en tiempo y espacio por una cantidad desconocida de generaciones.

Bibliografía

Anthony, M. 1954. Ecology of the Opuntiae in the big bend region of Texas. *Ecology* 35:334-347.

Arizaga, S. y E. Ezcurra. 1995. Insurance against reproductive failure in a semelparous plant: bulbil formation in *Agave macroacantha* flowering stalks. *Oecologia* 101:329-334.

Arizaga, S. y E. Ezcurra. 2002. Propagation mechanisms in *Agave macroacantha* (Agavaceae), a tropical arid-land succulent rosette. *American Journal of Botany* 89:632-641.

Ayres, D. R. y J. Ryan. 1997. The clonal and population structure of a rare endemic plant, *Wyethia reticulata* (Asteraceae): allozyme and RAPD analysis. *Molecular Ecology* 6:761-772.

Begon, M., J. L. Harper y C. Townsend. 1996. Ecology: individuals, populations and communities. Blackwell Science.

Bell, G. 1987. Two theories of sex and variation. En: S. C. Stearns [ed.]. *The Evolution of sex and its consequences*. Birkhäuser Verlag, Basilea.

Bierzichudek, P. 1982. The demography of jack-in-the-pulpit, a forest perennal that changes sex. *Ecological Monographs* 52:335-351.

Bush, S. P. y D. L. Mulcahy. 1999. The effects of regeneration by fragmentation upon clonal diversity in the tropical forest shrub *Pokilacanthus macranthus*: random amplified polymorphic DNA (RAPD) results. *Molecular Ecology* 8:865-870.

Carrillo Angeles, I. G. 2006. Distribución espacial de clones de *Ferocactus robustus*: efectos sobre la reproducción sexual. Tesis de Maestría. Posgrado en Ciencias Biológicas, Instituto de Ecología-UNAM.

Caswell, H. 1985. The evolutionary demography of clonal propagation. En: J. B. C. Jackson, L. W. Buss, y R. E. Cook [eds.]. *Population biology and evolution of clonal organisms*. Yale University Press, New Haven, Connecticut.

Caswell, H. 2001. *Matrix population models: construction, analysis, and interpretation*. Sinauer Associates Inc., Massachusetts.

Charlesworth, B. 1980. *Evolution in age-structured populations*. Cambridge University Press, Cambridge.

Charpentier, A. 2002 Consequences of clonal growht for plant mating. *Evolutionary Ecology* 15:521-530.

Clark-Tapia R., C. Alfonso-Corrado, L. E. Eguiarte y F. Molina-Freaner. 2005. Clonal diversity and distribution in *Stenocereus eruca* (Cactaceae), a narrow endemic cactus of the Sonoran Desert. *American Journal of Botany* 92:272-278.

Clark-Tapia, R., M.C. Mandujano, T. Valverde, A. Mendoza y F. Molina-Freaner. 2005. How important is clonal recruitment for population maintenance in rare plant species? : the case of of the narrow endemic cactus, *Stenocereus eruca*, in Baja California Sur, México. *Biological Conservation* 124:123-132.

Cochran, M. E., y S. Ellner. 1992. Simple methods for calculating age-based life history parameters for stage-structured populations. *Ecological Monographs* 62:345-364.

Cook, R. E. 1979. Asexual reproduction: a further consideration. *American Naturalist* 113:769-772.

Cook, R. E. 1985. Growht and development in clonal plant populations. En: J. B. C. Jackson, L. W. Buss, y R. E. Cook [eds.]. *Population biology and evolution of clonal organisms*. Yale University Press, New Haven, Connecticut.

de Kroon, H. y J. van Groenendael [eds.]. 1997. *The ecology and evolution of clonal plants*. Backhuys Publishers, Leiden.

de Kroon, H., A. Plaiser, J. M. van Groenendael y H. Caswell. 1986. Elasticity: the relative contribution of demographic parameters to population growth rate. *Ecology* 67:1427-1431.

de Steven, D. 1989. Genet and ramet demography of *Oenocarpus mapora* ssp. *mapora*, a clonal palm of panamanian tropical moist forest. *Journal of Ecology* 77:579-596.

Douhovinkoff, V. y R. S. Dodd. 2003. Intra-clonal variation and a similarity threshold for identification of clones: application to *Salix exigua* using AFLP molecular markers. *Theoretical Applied Genetics* 106:1307-1315.

Eguiarte, L. E., F. J. Nuñez, C. A. Domínguez y C. Cordero. 1999. Biología evolutiva de la reproducción de las plantas. En: J. Nuñez Farfán y L. E. Eguiarte [comp.]. *La evolución biológica*. UNAM, Conabio, México.

Ellstrand, N.C. y M. L. Roose. 1987. Patterns of genotypic diversity in clonal plant species. *American Journal of Botany* 74:123-131.

Elmqvist, T. y P. A. Cox. 1996. The evolution of vivipary in flowering plants. *Oikos* 77:3-9.

Eriksson, O. 1988. Ramet behaviour and population growth in the clonal herb *Potentilla anserina*. *Journal of Ecology* 76:522-536.

Eriksson, O. 1993. Dynamics of genets in clonal plants. *Trends in Ecology and Evolution* 8:313-316.

Falinska, K. 1995. Genet disintegration in *Filipendula ulmaria*: consequences for population dynamics and vegetation succesion. *Journal of Ecology* 83:9-21.

Flores-Martínez, A. 1994. Papel de *Mimosa luisana* en la estructura de la comunidad y su relación con *Neobuxbaumia tetetzo* en el Valle de Zapotitlán

de las Salinas, Puebla. Tesis de doctorado. UACEPyP, Centro de Ecología, UNAM.

Gabrielsen, T.M. y C. Brochmann. 1998. Sex after all: high levels of diversity detected in the arctic clonal plant *Saxifraga cernua* using RAPD markers. *Molecular Ecology* 7:1701-1708.

Gentry, H. S. 1998. *Agaves of continental North America*. The University of Arizona Press. Tucson.

Goldberg, E. D., y R. M. Turner. 1986. Vegetation change and plant demography in permanent plots in the Sonoran Desert. *Ecology* 67: 695-712.

Golubov, J., L.E. Eguiarte, M.C. Mandujano, J. López-Portillo, y C. Montaña. 1999. Why be a honeyless honey mesquite? Reproduction and mating system of nectarful and nectarless individuals. *American Journal of Botany* 86: 955-963.

Golubov, J., M.C. Mandujano, C. Montaña, J. López-Portillo y L. Eguiarte. 2004. The demographic costs of nectar production in the desert perennial *Prosopis glandulosa* (Mimosoideae): a modular approach. *Plant Ecology* 170:267-275.

Gotelli, N.J. 1991. Demographic models for *Leptogorgia virgulata*, a shallow water gorgonian. *Ecology* 72:457-467.

Gotelli, N.J. 1995. *A primer of Ecology*. Sinauer. Massachusets.

Grant, V. y K.A. Grant. 1980. Clonal microspecies of hybrid origin in the *Opuntia lindheimeri* group. *Botanical Gazzete* 141(1):101-106

Grime, J.P. 1982. Estrategias de Adaptación en las Plantas. Limusa. México D.F. 291 p.

Hadrys, H., Balick, M. y Scherwater, B. 1992. Applications of random amplified polymorphic DNA (RAPD) in molecular ecology. *Molecular Ecology* 1:55-63.

Hamrick, J.L., y M.J. Godt 1989. Allozyme diversity in plant species. En: A.H.D. Brown, M. T. Clegg, A. L. Kahler y B.S. Weir [eds.]. Plant population genetics, breeding, and genetic resources. Sinauer, Sunderland , Massachusets.

Handel, S.N. 1985. The intrusion of clonal growth patterns on plant breeding system. *American Naturalist* 125:367-384.

Harper, J. 1977. *Population biology of plants*. Academic Press, London.

Harper, J. 1985. Modules, branches, and the capture of resources. En: J. B. C. Jackson, L. W. Buss, y R. E. Cook [eds.]. *Population biology and evolution of clonal organisms*. Yale University Press, New Haven, Connecticut.

Hastings, A. 1996. *Population biology: concepts and models*. Springer-Verlag, New York.

Hicks, D. y A. Mauchamp. 1999. Population structure and growth patterns of *Opuntia echios* var. *gigantea* along an elevation gradient in the Galápagos Islands. *Biotropica* 32:235-243.

Horak, M.J., J.S. Holt y N.C. Ellstrand. 1987. Genetic variation in yellow nutsedge (*Cyperus esculentus*). *Weed Science* 35:506-512.

Huenneke, L.F., y P.L. Marks. 1987. Stem dynamics of the shrub *Alnus incana* sp. *rugosa*: transition matrix models. *Ecology* 68: 1234-1242.

Ivey, C.T. y J.H. Richards. 2001. Genotypic diversity and clonal structure of everglades sawgrass, *Cladium jamaicense* (Cyperaceae). *International Journal of Plant Science* 162(6):1327-1335.

Jackson, J.B.C., L.W. Buss y R.E. Cook. 1985. *Population biology and evolution of clonal organisms*. Yale University Press, New Haven, Connecticut.

Janzen, D. 1977. What are dandelions and aphids? *American Naturalist* 111:586-589.

Jonsson, B.O., Jónsdóttir, I.S. & Cronberg, N. 1996. Clonal diversity and allozyme variations in populations of the arctic sedge *Carex bigelowii* (Cyperaceae). *Journal of Ecology* 84:449-459.

Kanno, H., y Seiwa, K. 2004. Sexual vs. vegetative reproduction in relation to forest dynamics in the understorey shrub, *Hydrangea paniculata* (Saxifragaceae). *Plant Ecology* 170:43-53.

Krebs, Ch. 1985 *Ecología, estudio de la distribución y la abundancia*. Segunda edición. Harla Harper & Row Latinoamericana, México.

Kreher, S.A., S.A. Foré y Collins, B.S. 2000. Genetic variation within and among patches of the clonal species, *Vaccinium stamineum* L. *Molecular Ecology* 9:1247-1252.

Lefkovitch, L.P. 1965. The study of population growth in organisms grouped by stages. *Biometrics* 21:1-18.

Liston, A., B.L. Wilson, W.A. Robinson, P.S. Doescher, N.R. Harris y T. Svejear. 2003. The relative importance of sexual reproduction versus clonal spread in aridland bunchgrass. *Oecologia* 137:216-225.

López-Portillo, J., C. Montaña y E. Ezcurra. 1996. Stem demography of *Prosopis glandulosa* var. *torreyana* in vegetation arcs and associated bare areas. *Journal of Vegetation Science* 7:901-910.

MacArthur, R.H. 1984. Geographical Ecology: Patterns in the Distribution of species. Princeton Universtiy Press. USA.

Mandujano, M.C. 1995. Establecimiento por semilla y propagación vegetativa de *Opuntia rastrera* en la Reserva de la Biósfera de Mapimí, Durango. Tesis Doctoral, Universidad Nacional Autónoma de México, México.

Mandujano, M.C., C. Montaña, y L.E. Eguiarte. 1996. Reproductive ecology and inbreeding depression in *Opuntia rastrera* (Cactaceae) in the Chihuahuan desert: why are sexually derived recruitments so rare?. *American Journal of Botany* 83:63-70.

Mandujano, M.C., J. Golubov y C. Montaña. 1997. Dormancy and endozoochorous dispersal of *Opuntia rastrera* seeds in the Chihuahuan desert. *Journal of Arid Environments* 36:259-266.

Mandujano, M.C., C. Montaña, I. Méndez y J. Golubov. 1998. The relative contributions of sexual reproduction and clonal propagation in *Opuntia rastrera* from two habitats in the Chihuahuan Desert. *Journal of Ecology* 86:911-921.

Mandujano, M., Montaña, C., Franco, M., Golubov, J. y Flores-Martínez, A. 2001. Integration of demographic annual variablity in a clonal desert cactus. *Ecology* 82:344-359.

Mesterton-Gibbons, M. 1993. Why demographic elsaticities sum to one: a postscript to de Kroon *et al*. *Ecology* 74:2467-2468.

Mogie, M.y M. Hutchings. 1990. *Phylogeny, ontogeny and clonal growth in vascular plants*. SPB Academic Publishing, The Hague.

Molina Freaner, F. y R. Clark Tapia. 2005. Clonal diversity and allelic relationships between two closely related species of columnar cacti form the Sonoran Desert: *Stenocereus eruca* and *Stenocereus gummosus*. *International Journal of Plant Science*. 166(2): 505-511.

Nault, A. y D. Gagnon. Ramet demography of *Allium tricoccum,* a spring ephemeral, perennial forest herb. *Journal of Ecology* 81:101-119.

Nei, M. y W.-H. Li. 1979. Mathematical model for studying genetic variation in terms of restriction endonucleases. *Proceedings of the National Academy of Sciences*, USA 76:5269-5273.

Palleiro, N., Mandujano, M.C. y J. Golubov. 2006. Aborted fruits of *Opuntia microdasys* (Cactaceae): insurance against reproductive failure. *American Journal of Botany* 93(4). En prensa.

Pappert, R.A., J. L. Hamrick y L.A. Donovan. 2000. Genetic variation in *Pueraria lobata* (Fabaceae), an introduced, clonal, invasive plant of the Southeastern United States. *American Journal of Botany* 87:1240-1245.

Pielou, E.C. 1969. *An introduction to mathematical ecology*. Wiley, Nueva York.

Pimienta-Barrios, E. y R.F. del Castillo. 2002. Reproductive Biology. En: P.S. Nobel [ed.]. *Cacti: biology and uses*. University of California Press. Berkeley.

Resnick, R.y D. Halliday. *1992. Basic concepts in relativity.* Macmillan. New York.

Reusch, T.B.H., W.T. Stam y J.L. Olsen. 1998. Size and estimated age of genets in eelgrass, *Zostera marina*, assessed with microsatellite markers. *Marine Biology* 133:519-525.

Richards, C.L., J.L. Hamrick, L.A. Donovan y R. Mauricio. 2004. Unexpectedly high clonal diversity of two salt marsh perennials across a severe environmental gradient. *Ecology Letters* 7 (12):1155-1162.

Shimizu, Y., M. Ando y F. Sakai. 2002. Clonal structure of natural populations of *Cryptomeria japonica* growing at different positions on slopes, detected using RAPD markers. *Biochemical Systematics and Ecology* 30:733-748.

Silander, J.A. 1984. The genetic basis of the ecological amplitude of *Spartina patens*. III. Allozyme variation. *Botanical Gazette* 145:569-577.

Silva, A. y L.E. Eguiarte. 2003. Geographic patterns in the reproductive ecology of *Agave lechuguilla* (Agavaceae) in the Chihuahuan Desert. II. Genetic variation, differentiation, and inbreeding estimates. *American Journal of Botany* 90:700-706.

Silvertown, J., M. Franco, I. Pisanty y A. Mendoza. 1993. Comparative plant demography-relative importance of life-cycle components to the finite rate of increase in woody and herbaceous perennials. *Journal of Ecology* 81:465-476.

Silvertown, J., y J. Lovett-Doust. 1993. *Introduction to plant population biology*. Blackwell Scientific Publications, Cambridge.

Stehlik, I. y R. Holderegger. 2000. Spatial genetic structure and clonal diversity of *Anemone nemorosa* in late successional deciduous woodlands of Central Europe. *Journal of Ecology* 88:424-435.

Steinger, T., C. Körner y B. Schmid. 1996. Long-term persistence in a changing climate: DNA analysis suggest very old ages of clones of alpine *Carex curvula*. *Oecologia* 105:94-99.

Sydes, M.A. y R. Peakall. 1998. Extensive clonality in the endangered shrub *Haloragodendron lucasii* (Haloragaceae) revealed by allozymes and RAPDs. *Molecular Ecology* 7:87-93.

Tiffney, B.H. y K.J. Niklas. 1985. *Clonal growth in land plants: a paleobotanical perspective*. En: J.B.C. Jackson, L. Buss y R. E. Cook [eds.]. *Population biology and evolution of clonal organisms*. Yale University Press.

van Groenendael, J., H. de Kroon y H. Caswell. 1988. Projection matrices in population biology. *Trends in Ecology and Evolution* 3(10):264-269.

van Groenendael, J., H. de Kroon, S. Káliz y S. Tuljapurkar. 1994. Loop analysis: evaluating life history pathways in population projection matrices. *Ecology* 75:2410-2415.

Vasek, F. C. 1980. Cerosote bush: long-lived clones in the Mojave Desert. *American Journal of Botany* 67:246-255.

Vavrek, M. C., J. B. Mc Graw y H.S. Yang. 1997. Within-population variation in demography of *Taraxacum officinale*: season- and size-dependent survival, growth and reproduction. *Journal of Ecology* 85:277-287.

Watson, M.A. 1984. Developmental constraints: effect on population growth and patterns of resource allocation in a clonal plant. *The American Naturalist* 123:411-426.

Wijesinghe, D.K. 1994. Temporal and structural components of ramet independence in the clonal perennial herb, *Potentilla simplex*. *Journal of Ecology* 82: 13-20.

Williams, G.C. 1975. Sex and evolution. Princeton, New Jersey.

Willson, M.F. 1979. Sexual selection in plants. *The American Naturalist* 113:777-790.

Willson, M.F. 1983. *Plant reproductive ecology*. John Wiley & Sons.

Wolf, A., S.P. Harrison y J.L. Hamrick. 1999. Influence of habitat patchiness on genetic diversity and spatial structure of a serpentine endemic plant. *Conservation Biology* 14:454-463.

Wolfe, A.D., Q.-Y. Siang y S. Kephart. 1998. Assessing hybridization in natural populations of *Pentetemon* (Scrofulariaceae) using hypervariable intersimple sequence repeat (ISSR) bands. *Molecular Ecology* 7:1107-1125.

Wyman, J., A. Bruneau y M.F. Tremblay. 2003. Microsatellite analysis of genetic diversity in four populations of *Populus tremuloides* in Quebec. *Canadian Journal of Botany* 81:360-367.

Capítulo 8

Ecología molecular de la conservación

Martha Rocha y Jaime Gasca

Actualmente la diversidad de los seres vivos atraviesa por una crisis, ya que muchas especies se encuentran amenazadas o en peligro y en el peor de los casos, ya se han extinguido por efecto de las actividades del hombre. En los últimos años se ha hecho una infinidad de estudios para comprender los procesos ecológicos y evolutivos por los que atraviesan las especies y tratar de proponer estrategias de conservación que sean exitosas. Sin embargo, aún nos encontramos en la fase de construcción del cuerpo teórico que nos permitirá preservar de la mejor manera las especies, y se ha sugerido que se requiere de un trabajo "transdisciplinario", donde se involucren profesionales tanto de las ciencias naturales como sociales. Dentro de las ciencias naturales se requieren especialistas de diversas disciplinas como la taxonomía, la ecología, la botánica, la zoología y la biología evolutiva. En este capítulo nos enfocaremos en el papel de los biólogos evolutivos, en particular de los ecólogos moleculares, y qué pueden éstos aportar a la teoría sobre la conservación de especies.

La genética de poblaciones es una parte central de la teoría evolutiva moderna, y sus aportaciones a la biología de la conservación han ido creciendo conforme su teoría y práctica han ido integrándose en la disciplina que ahora se conoce como Genética de la Conservación (Eguiarte y Piñero, 1990).

Un poco de historia

Los orígenes de la genética de la conservación se dan poco después de haber surgido la biología de la conservación, cuando se hicieron evidentes varios problemas genéticos asociados con las especies en peligro de extinción. Por ejemplo, se resaltó que la disminución de los tamaños poblacionales iba acompañada de la pérdida de diversidad genética y que la fragmentación de los hábitats tenía un efecto en la estructura poblacional. Desde el punto de vista evolutivo, se hizo entonces necesario entender los procesos de extinción de las especies (Simberloff, 1988, Eguiarte y Piñero, 1990).

Sobre las bases de la teoría de biogeografía de islas, en los años 70, se desarrolló la teoría para el diseño de refugios y para la toma de decisiones al planear los tamaños y formas óptimas de las reservas, así como la conectividad entre ellas. En la década de los 80 comenzaron a usarse sistemáticamente los análisis genéticos en especies en cautiverio, se realizaron los primeros experimentos en laboratorio y se avanzó en los estudios de metapoblaciones, enfocados en la cuantificación de la variación genética y los tamaños poblacionales efectivos. El avance en el uso de marcadores moleculares permitió a su vez una espectacular recopilación de datos genéticos de las poblaciones naturales de especies amenazadas o en peligro y mostró la relevancia de los factores genéticos, particularmente en casos como la depresión por endogamia (Meffe y Carroll, 1994; Eguiarte y Piñero, 1990; Primack *et al.*, 2001).

En la década de los 90 los principales avances estuvieron relacionados con los nuevos métodos moleculares y computacionales que permitieron nuevos análisis y predicciones, dando mayor importancia al reconocimiento de las unidades fundamentales de conservación (especies, subespecies y ESU o *Unidades Evolutivamente Significativas* -ver más adelante).

El rápido avance de la genética de la conservación se puso de manifiesto con la publicación en el año 2000 de la revista *Conservation Genetics* (http://www.kluweronline.com/jrnltoc.htm/1566-0621) y la reciente aparición de libros como *Introduction to conservation genetics* (Frankham et al., 2002), así como la sección *Population and conservation genetics* en la revista *Molecular Ecology*. (http://www.blackwelloublishing.com/journals/mec).

Pero, ¿qué relación guardan la genética y la conservación?

Uno de los objetivos clave de la genética de la conservación es ayudar a minimizar las extinciones evitando los problemas relacionados con tamaños efectivos pequeños (véase el capítulo 3 de este libro), como el efecto deletéreo de la endogamia (depresión por endogamia; véase el capítulo 6 de este libro), la pérdida de diversidad y la habilidad para evolucionar en respuesta a los cambios ambientales, así como los efectos deletéreos que ocurren por la cruza entre individuos muy distintos (depresión por exogamia; Amos y Balmford, 2001; Frankham *et al.*, 2002). Los análisis genéticos también permiten estudiar el efecto de la fragmentación y la reducción del flujo génico en poblaciones estructuradas y el efecto de la acumulación y pérdida de mutaciones deletéreas. Por otro lado, con la genética también se puede dar solución a problemas de tipo taxonómico, por ejemplo establecer especies prioritarias para la conservación, resolver incertidumbres taxonómicas, definir unidades evolutivamente significativas y unidades de manejo, y en el ideal, poder proteger los procesos evolutivos que mantienen la diversidad biológica (Moritz, 2002).

Una aplicación más de la genética de la conservación en la práctica es la identificación de material biológico y su procedencia, ya que muchas veces resulta muy difícil identificar si un organismo o producto (huevos, carne, huesos, semillas, etc.) proviene de especies protegidas por la ley. Existen varios ejemplos que han demostrado la efectividad de las técnicas moleculares para detener el comercio ilegal de recursos biológicos. Un caso particular ha sido el de las ballenas, grupo en el cual muchas especies se encuentran amenazadas y por tanto protegidas por la ley, y a pesar de esto se continúa la cacería para consumo de su carne. Baker y colaboradores (2000) desarrollaron métodos con ADN mitocondrial para identificar la especie a la que pertenecían muestras de carne de ballena, y encontraron que un 9% de la carne de mercados de Japón y Korea proviene de especies amenazadas. Para el caso de los tiburones (cuya aleta dorsal es muy apreciada en Asia), Shivji y colaboradores (2002) diseñaron marcadores ribosomales específicos para varias especies, lo que facilitará el reconocimiento de productos que provengan de tiburones protegidos. Las técnicas anteriores no sólo sirven en análisis forenses, ya que también se han usado para determinar la procedencia de organismos que han sido extraídos de su hábitat para su posterior reintroducción (Frankham *et al.*, 2002).

Especies en peligro, ¿a quién hay que proteger?

En principio, es necesario establecer los criterios para decidir si una especie debe o no ser conservada, o más precisamente, si se encuentra en un estado crítico donde la conservación sea de carácter urgente. La Unión Internacional para la Conservación de la Naturaleza (IUCN, *International Union for Conservation of Nature and Natural Resources*) ha determinado tres categorías: "críticamente en peligro", "peligro" y "vulnerable", en donde se incluyen taxa que se consideran en amenaza de extinción. Tales categorías se basan en tres criterios principales: i) la **disminución** de las poblaciones, que se refiere a una reducción del tamaño de la población en un 90%, 70% y 50%, para cada categoría respectivamente, en los últimos 10 años o las últimas tres generaciones; ii) el **área de ocupación**, que propone áreas de de 100 km^2, 500 km^2 y 2 000 km^2, para cada categoría respectivamente y iii) los tamaños poblacionales, que se miden en número de individuos maduros y son de 50, 2 500 y 10 000 para cada categoría respectivamente (UICN, 2001). Esta misma unión reconoce tres niveles en que debe conservarse la naturaleza: la diversidad genética, la diversidad de especies y la diversidad de ecosistemas (McNeely *et al.,* 1990).

Como podemos ver, los análisis genéticos son fundamentales para estas caracterizaciones, ya que nos permiten calcular tanto los tamaños efectivos de las poblaciones (cabe señalar, sin embargo, que según el criterio iii) la UICN reconoce tamaños poblacionales censales y no tamaños poblacionales efectivos, lo que sería más adecuado), además de la diversidad genética, aunque veremos más adelante que hay muchos otros factores genéticos importantes que deben ser tomados en cuenta.

Principios básicos en genética de la conservación

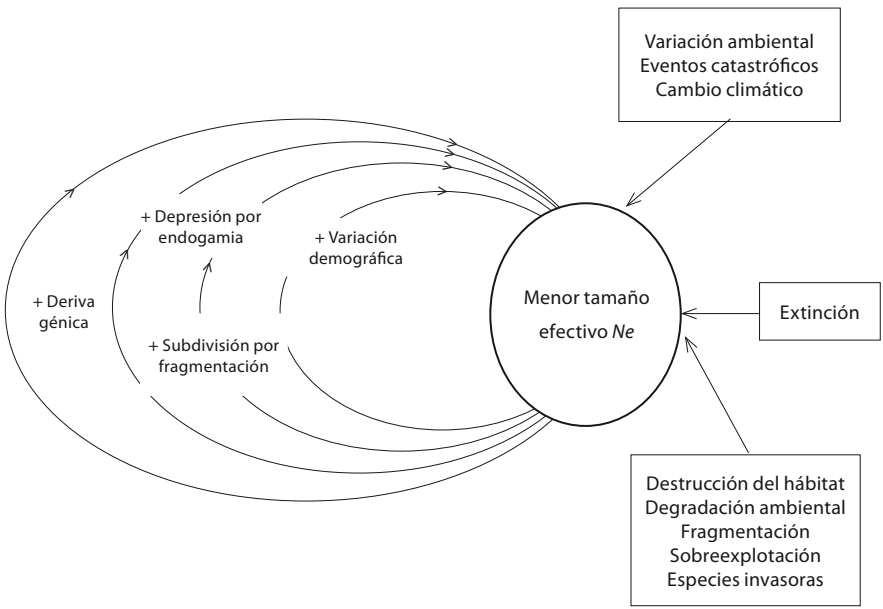
Causas genéticas de la extinción, problemas de las poblaciones pequeñas

El objetivo central de la genética de la conservación en los últimos años ha sido entender y disminuir los problemas genéticos enfrentados por las poblaciones pequeñas. A la fecha existen numerosos ejemplos donde los factores genéticos parecen estar implicados en la disminución de poblaciones de animales como las panteras de Florida (*Felis concolor*), los pericos de Puerto Rico (*Amazona vittata*), los lobos de la isla Royal (*Canis lupus*), el borrego cimarrón (*Ovis*

canadensis), el pájaro carpintero *(Dendrocopos medius)* y los leones asiáticos (*Panthera leo persica*), entre otros (Hedrick ,1995; O'Brien, 1994; Frankham *et al.*, 2002). Entre los factores genéticos que se han reconocido, la pérdida de variación genética y la depresión por endogamia han recibido la mayor atención, por lo que serán tratadas con mayor cuidado más adelante.

Cuando las poblaciones son pequeñas son más propensas a la extinción, ya que los factores estocásticos (tanto genéticos como demográficos, ambientales y catástrofes) aceleran su decline (Primack *et al.*, 2001) y las llevan a los llamados **vórtices de extinción** (Primack *et al.*, 2001; Frankham *et al.*, 2002). Éstos se inician cuando las poblaciones son pequeñas, demográficamente inestables y tienen niveles altos de endogamia; la disminución poblacional reduce la adecuación y se produce una retroalimentación negativa, lo que refuerza el decline, aumenta la depresión por endogamia y la susceptibilidad a eventos estocásticos reduciendo aún más el tamaño poblacional, y así sucesivamente hasta llegar a la extinción (figura 1).

Figura 1. Cuando una población es muy pequeña, los fenómenos estocásticos incrementan el efecto de los problemas genéticos, por lo que la población sufre un vórtice de extinción



Otro problema que se ha descrito es el llamado **colapso mutacional** (o *mutational meltdown*). Lynch (1993) observó que las mutaciones deletéreas se acumulan más en poblaciones pequeñas, ya que la selección es menos eficiente. Cuando las poblaciones son grandes las mutaciones deletéreas son removidas por la selección natural y ésta elimina a los individuos con mayor carga genética. Si la población comienza a declinar, el número de individuos disponible para que se eliminen las mutaciones deletéreas disminuye, la purga se vuelve cada vez menos eficiente, aumenta la carga genética de la población y la adecuación disminuye (Lynch, 1993; Lande, 1995; Lynch *et al.*, 1995). El resultado es una retroalimentación negativa similar al vórtice de extinción (Amos y Balmford, 2001).

Los **cuellos de botella**, por otra parte, son reducciones drásticas en los tamaños efectivos y pueden repercutir en los niveles de variación genética. Si las poblaciones permanecen pequeñas por largos periodos de tiempo, el efecto de error de muestreo es acumulativo. Esto genera cambios al azar en las frecuencias alélicas, lo que se conoce como deriva génica (Hartl y Clark, 1997). Debido a que en poblaciones de mayor tamaño las fluctuaciones no son tan grandes, se espera que mantengan niveles de variación genética mayores que en las poblaciones pequeñas.

Un ejemplo clásico del efecto de la deriva génica después de un cuello de botella es el caso del elefante marino *Mirounga angustirostris*, que fue sujeto a intensa cacería durante el siglo XIX, hasta casi llegar a la extinción. En 1892 fueron reportados 8 individuos (de los cuales mataron a 7) en la isla de Guadalupe; para 1922 lentamente se iba recuperando una población relictual que contaba ya con 350 individuos que además comenzaron a recibir protección legal y, para los años 80 se estimó un número de entre 550,000 y 750,000 individuos. Los análisis genéticos mostraron que la población tuvo un 20% de probabilidad de extinción y que poseía un monomorfismo genético en isoenzimas debido al severo cuello de botella. Actualmente se cree que el haber pasado por este cuello de botella puede tener repercusiones como la vulnerabilidad a agentes infecciosos y la presencia de problemas en la reproducción (Hoelzel *et al.*, 1993).

Para poder entender mejor las consecuencias de los tamaños pequeños, es necesario identificar si la población ha sufrido reducciones drásticas recientes en su tamaño, para lo que se han desarrollado métodos estadísticos que infieren los cambios en el pasado y modelos que ayudan a distinguir las razones ecológicas de la reducción (para una revisión ver Beaumont, 2001).

Causas ecológicas de la extinción: controversias

Cuando comenzaron a acumularse evidencias de las causas genéticas de la extinción, los estudios genéticos comenzaron a cobrar importancia en los programas de conservación. Sin embargo, algunos expertos en el área se dieron cuenta de que las poblaciones se enfrentaban a otros problemas de más corto plazo, como las fluctuaciones de los tamaños poblacionales, el reclutamiento y la sobrevivencia; mientras que la pérdida de diversidad genética o la depresión por endogamia parecían menos tangibles y tener efecto más bien a largo plazo (Moritz *et al.*, 1993).

En 1988, Lande publicó un artículo muy persuasivo en el que concluyó que la estocasticidad ambiental y demográfica pueden llevar a las poblaciones pequeñas a la extinción, antes de que los factores genéticos desempeñen un papel importante. Trabajos posteriores interpretaron esta crítica como si no se debiera confiar en los datos genéticos y solamente se debieran emplear los datos demográficos (Caughley, 1994) o bien ser muy cuidadosos con los resultados obtenidos de la genética (Caro y Laurenson, 1994). Sin embargo, la crítica de Lande era en el sentido de que no se emplearan solamente datos genéticos sino también otros aspectos de la ecología de las poblaciones. Así lo demuestran Westemeier y colaboradores (1998), quienes encontraron cambios demográficos y genéticos en el urogallo (*Tympanuchus cupido*) durante 35 años y mostraron evidencias de que había una interacción de la genética con la pérdida de hábitat y la variación estocástica, lo que daba como resultado una disminución de los tamaños poblacionales de manera similar al vórtice de extinción.

Mills y Smouse (1994) mostraron que los estudios de viabilidad poblacional no necesariamente deben marcar una dicotomía entre los factores genéticos y las fluctuaciones demográficas y ambientales, sino que tanto la estocasticidad genética (endogamia) como las fuerzas estocásticas no genéticas son sumamente importantes para determinar la probabilidad de extinción. Estos autores hacen énfasis en que tanto la genética como la ecología deben ser tomadas en cuenta, así como las interacciones entre ellas.

La pérdida de diversidad genética

La diversidad genética es sin lugar a dudas la materia prima para la evolución, ya que de ella dependen tanto la adaptación como la especiación. Los niveles altos de diversidad pueden dar la habilidad para responder a enfermedades,

parásitos, depredadores y cambios ambientales (Hedrick, 2001), y fue por ello que los primeros trabajos de conservación buscaban simplemente mantener niveles de diversidad altos.

En general se han encontrado patrones en los niveles de diversidad con los que se hacen inferencias en los estudios de conservación; por ejemplo, las poblaciones grandes, que se entrecruzan, tienen altos niveles de diversidad genética, mientras que las poblaciones pequeñas y asexuales tienen niveles de diversidad bajos (Amos y Harwood, 1998).

No obstante, no siempre se cumplen las predicciones evolutivas con los patrones de diversidad observados: algunas especies para las que se predice poca variación tienen niveles altos, mientras que otras especies presentan mucha menor variación de la esperada. Al parecer, no se puede generalizar la regla de que altos niveles de diversidad significan buen estado evolutivo. Por ejemplo, se encuentran especies muy abundantes e incluso invasivas que tienen poca variación (como las hierbas acuáticas y algunos pastos) y especies con mucha variación y cuyas poblaciones están declinando (Baur y Schmid, 1996; Amos y Harwood, 1998).

Por otro lado, generalmente se asume que la reducción en la diversidad se correlaciona con una baja adecuación, los ejemplos clásicos son la asimetría en los dientes del chita (*Acinonyx jubatus*), los defectos anatómicos en el lince (*Linx pardinus*) o el bajo reclutamiento en el trébol *Trifolium hirtum (*Baur y Schmid, 1996).

Entre los muchos ejemplos de organismos en que se ha cuantificado la diversidad genética, mencionaremos el caso de la nutria marina (*Enhydra lutris*), cuyas poblaciones han pasado por cuellos de botella debido al comercio de pieles y tienen niveles bajos de variación. Tal carencia de variación pudo ser resultado de la extensiva cacería, o bien a niveles de variación históricamente bajos. Para comprobar esto, Larson y colaboradores (2002) analizaron marcadores mitocondriales en huesos de nutrias que vivieron antes del comercio de pieles y los compararon con datos de las poblaciones actuales. Los resultados mostraron una reducción de al menos un 62% de los alelos y un 43% de heterocigosis en las poblaciones actuales, lo que apoyó la idea de la reducción debida a la cacería. Los datos obtenidos sugieren que debe llevarse a cabo un monitoreo continuo de las poblaciones para determinar si la pérdida de variación genética pudiera tener un efecto en la sobrevivencia a largo plazo de las nutrias marinas.

El paradigma de la baja diversidad en especies raras o amenazadas

Analizar los niveles de diversidad en plantas resulta complejo por diversas razones, entre otras por que pueden presentar sistemas complejos de polinización y de dispersión (Falk y Holsinger, 1991). Recientemente se han usado los patrones de distribución espacial para definir a las especies "raras" como aquéllas que presentan una distribución espacial restringida y en éstas enfocar los esfuerzos de conservación (Rabinowitz, 1981; Kunin y Gaston, 1993).

Se ha observado que la distribución y niveles de variación genética en las poblaciones de especies raras son muy variables, y dependen de si la especie ha sido históricamente rara o la rareza ha sido reciente, en tal caso, muy probablemente por causas humanas (Bevill y Louda, 1999). Debido a la complejidad en el análisis de la diversidad en especies raras, una estrategia que se ha seguido es comparar especies raras y comunes que pertenezcan al mismo género. Los primeros trabajos indicaron que las especies raras tienen niveles de variación significativamente más bajos que sus congéneres con amplia distribución (Karron, 1987; Frankham, 1995); sin embargo, también hay casos en los que la variación es sorprendentemente alta para especies raras y, por otro lado, en algunos casos ambas especies (raras y comunes) tienen niveles muy altos o muy bajos de variación (Young y Brown, 1996).

Gitzendanner y Soltis (2000) publicaron una revisión de todos los trabajos reportados donde se comparan especies raras y comunes y también llegaron a conclusiones contradictorias: por una parte, el porcentaje de loci polimórficos, el número de alelos por locus y la heterocigosis fueron menores en especies raras que en las comunes; sin embargo, entre el 24% y el 29% de las especies raras analizadas fueron tanto o más variables que las comunes. Posteriormente, Cole (2003) amplió la comparación de congéneres raros y comunes a un mayor número de especies y concluyó al igual que Gitzendanner y Soltis que los índices de variación son menores en especies raras, aunque la diferencia en los sistemas reproductivos afecta la manera en que se estructura la diversidad, de modo que se pone de manifiesto que los niveles de variación también están determinados por atributos ecológicos. Por lo anterior, el alto porcentaje de especies raras con igual o mayor variación que las especies comunes hace necesaria la evaluación de otros factores como los atributos reproductivos, las formas de vida, la demografía, las estrategias del ciclo de vida, los caracteres de respuesta al ambiente y las interacciones con

otros organismos, ya que todos estos factores pueden afectar los niveles de variación genética (Bevill y Louda, 1999).

Actualmente las prácticas de manejo están generalmente centradas en mantener altos niveles de diversidad y no consideran que en algunos casos puede ser normal que existan niveles bajos por razones históricas o ecológicas.

Diversidad y sistemas de apareamiento

Tanto en plantas como en animales la estructura y diversidad genética refleja en gran medida el sistema reproductivo de la especie. La variación genética es mínima en individuos de una misma clona y la variación se mide como el número de individuos por clona (Baur y Schmid, 1996); por ejemplo, se ha encontrado falta de variación en isoenzimas en gasterópodos terrestres que se auto-fertilizan (Baur y Klemm, 1989). Un trabajo interesante sobre los niveles de diversidad se hizo con el género *Styolanthes: S. humilis* es una planta anual, mientras que *S. viscosa* es perenne. Sawkins y colaboradores (2001) encontraron que *S. humilis* tiene menor variación genética que *S. viscosa*, lo cual puede deberse a la diferencia en los sistemas reproductivos. Existen varias revisiones a este respecto, por ejemplo, Loveless y Hamrick (1984) y Hamrick y Godt (1990) compararon atributos de la historia de vida con la estructura genética en plantas. Ellos observaron que las plantas que presentan tiempos generacionales cortos, que eran polinizadas por animales, que tenían baja fecundidad y con baja dispersión de semillas, presentaban a su vez poco flujo génico, tamaños efectivos reducidos y baja diversidad. En otro trabajo clásico, pero con animales, Nevo (1983) encontró que el tamaño efectivo, la fecundidad y la amplitud de nicho estaban positivamente asociados con la diversidad genética.

Una reflexión importante sobre los niveles de diversidad genética es que éstos dependen del sistema reproductivo y de la historia de la especie; por ejemplo, es posible que parte de las especies que están enlistadas como amenazadas hayan sido siempre raras y hayan persistido en poblaciones pequeñas durante su historia, mientras que otras hayan sufrido recientemente una importante reducción en sus poblaciones, que eran anteriormente grandes y ahora deben ser prioritarias para la conservación (Moritz *et al.*, 1993).

Restauración de la diversidad genética

Se sabe que los inmigrantes de poblaciones aledañas pueden prevenir la extinción de poblaciones pequeñas por el proceso llamado *rescate genético*

(Ingvarsson, 2001) o *restauración genética* (Hedrick, 2005). En mariposas, aves y plantas se ha visto que las cruzas entre diferentes poblaciones generan una **heterosis** (también llamado vigor híbrido) en donde la heterocigosis *per se* tiene una influencia positiva sobre la adecuación y además disminuye la expresión de los alelos recesivos deletéreos. Por esto es importante mantener el flujo génico entre las poblaciones que quedan en hábitats fragmentados para mantener aquellas que son sensibles a la endogamia (Keller y Waller, 2002).

En estudios experimentales con *Drosophila* se mostró que la adición de un solo inmigrante en una población endógama puede tener un efecto impresionante y mejorar la adecuación hasta en un 50% (Spielman y Frankham, 1992). Recientemente se publicó un ejemplo muy interesante de rescate genético en el lobo gris escandinavo (*Canis lupus*), en donde la inmigración de un solo individuo provocó el aumento en la heterocigosis, se evitó la endogamia, se dispersaron rápidamente los nuevos alelos en la población y aumentó la tasa de crecimiento poblacional (Vilà *et al.*, 2003).

Tamaños efectivos y tamaño mínimo viable (TMV)

Dado que los recursos tanto físicos como monetarios necesarios para la conservación de especies son enormes, resulta necesario tomar medidas de la manera más pragmática posible, ya que por una parte la retención de pocos individuos puede tener efectos genéticos deletéreos, pero también puede resultar contraproducente destinar demasiados recursos a una sola especie. En este sentido, un aspecto importante en la caracterización genética de las especies en peligro es poder calcular los tamaños efectivos (N_e) de las poblaciones. La metodología para estimar N_e ha sido revisada por Moreno (capítulo 3 de este libro) y también puede verse la revisión de Beaumont (2001).

En el ámbito de la conservación se ha llegado a la definición de tamaño mínimo viable (TMV) como "el número de individuos necesarios para asegurar la supervivencia de una especie a largo plazo"; es el tamaño poblacional más pequeño que tiene un alta probabilidad de sobrevivir en el futuro cercano.

Franklin (1980) sugiere que un tamaño efectivo de 500 sería necesario para mantener los niveles de diversidad encontrados en la naturaleza, basándose en un balance entre deriva y mutación. Soulé y colaboradores (1986) sugieren que el principal objetivo es mantener el 90% de la variación en 200 años. El tamaño estaría en función del tamaño de la población fundadora, la tasa de crecimiento poblacional y el tiempo generacional.

Debido al estado actual de amenaza de muchas especies, el desarrollo del TMV resulta vital para la recuperación de sus poblaciones. Diversos trabajos han calculado los TMV para diversas poblaciones, como las de rinocerontes de un cuerno de la India (*Rhinoceros unicornis*), que a pesar de mostrar niveles normales de variación genética en isoenzimas (Dinerstein y McCracken, 1990), se han presentado argumentos que anticipan un lento deterioro genético a largo plazo (Frankham, *et al.*, 2002). Otro ejemplo es el de los leones dorados tamarinos (*Leontopithecus rosalie*) en donde se calculó un TMV de 400 individuos para mantener el 98% de su diversidad genética dentro de los próximos 100 años (Ballou *et al.*, 1998).

Variación en caracteres cuantitativos

Hasta el momento nos hemos referido a los niveles de diversidad genética según los datos de marcadores de loci individuales, como son los marcadores de ADN o las isoenzimas. Por otro lado, existe la genética cuantitativa, que es un área de investigación muy amplia que analiza caracteres poligénicos a partir de valores fenotípicos que se distribuyen de manera continua, es decir, caracteres cuantitativos (Falconer y Mackay, 1996; Cap. 11 en Hedrick, 2000). La preocupación de los genetistas cuantitativos al hacer análisis de conservación con técnicas de loci sencillos es el estimar si la variación está correlacionada con la variación en caracteres ecológicos importantes que le confieran al individuo la capacidad de sobrevivir a cambios del ambiente (Storfer, 1996). Se ha visto que la diversidad que es resultado de la evolución adaptativa se mide preferentemente con la variación fenotípica, ya que en pocas ocasiones los análisis moleculares con pocos genes pueden predecir la variación en la adecuación total.

Reed y Frankham (2001), con un meta-análisis, demostraron que las medidas de marcadores moleculares sólo explicaban un 4% de la varianza en caracteres cuantitativos y lo que más llamó su atención era que la relación era aún menor para caracteres de interés en conservación, como aquellos relacionados con la historia de vida. Esto se explica porque los marcadores neutros evolucionan principalmente por deriva, mientras que los caracteres cuantitativos lo hacen por selección, ya que están directamente relacionados con la sobrevivencia (Reed y Frankham, 2001; Frankham *et al.*, 2002).

La heterocigosis en genes que afectan caracteres cuantitativos es proporcional a la varianza genotípica que está determinada genéticamente (y no por el ambiente), y el parámetro que mide la importancia de esta varianza es llamado

heredabilidad (Falconer y Mackay, 1996; Hedrick, 2000). Cuando la varianza genotípica es muy baja, el riesgo de extinción es más alto por la incapacidad de responder a los cambios del medio ambiente. Estimar la heredabilidad puede servir para indicar la respuesta a la selección y a su vez el potencial evolutivo de las poblaciones, ya que la correlación entre heredabilidad y adecuación es muy alta (Reed y Frankham, 2003).

Por otro lado se ha observado que los caracteres cuantitativos retienen mas variación en poblaciones pequeñas y se recuperan más rápidamente de los cuellos de botella que las isoenzimas (Lande y Barrowclough, 1987). Asimismo, existen muchos ejemplos en los que las especies exhiben niveles de diversidad fenotípica muy alta y habitan en ambientes extremadamente variables, aunque muestren niveles de diversidad isoenzimática bajos (Reed y Frankham, 2003).

Depresión por endogamia

El mismo Darwin reconoció la disminución de la adecuación cuando hay endogamia. Durante muchos años analizó el efecto de cruzas en plantas y se dio cuenta de que la entrecruza era generalmente benéfica y la autocruza perjudicial (Darwin, 1876). La endogamia, como sabemos, es una fuerza evolutiva que actúa cuando los apareamientos no son al azar sino entre individuos relacionados. En este caso los alelos idénticos por descendencia se encuentran juntos con mayor frecuencia y se redistribuyen las frecuencias de los genotipos aumentando los homócigos. Las mutaciones deletéreas que normalmente se van acumulando de manera recesiva aumentan (provocando una **carga genética** para la población), teniendo como resultado una reducción en la adecuación que se conoce como "depresión por endogamia" (DE; Hedrick, 2000; véase el capítulo 6 de este libro).

Existen dos explicaciones genéticas alternativas a la depresión por endogamia; una es la dominancia, en donde la adecuación baja por la homocigosis en loci particulares con alelos recesivos deletéreos y que se encuentran enmascarados en las poblaciones no endógamas por los alelos dominantes no deletéreos (Husband y Schemske, 1996; Byers y Waller, 1999; Charlesworth y Charlesworth, 1999; Frankham *et al.*, 2002). Por otro lado, está la muy difundida idea de la sobredominancia o vigor híbrido, en que la heterocigosis *per se* tiene una influencia positiva sobre la adecuación. Charlesworth y Charlesworth (1999) demuestran con experimentos en *Drosophila* y con evidencias en plantas, que la hipótesis del efecto de los alelos recesivos, algunos mantenidos

por mutación y otros por selección balanceadora (por ventaja del heterócigo), parecen ser la principal causa de depresión por endogamia.

Un ejemplo de la repercusión que puede tener la DE se observó en *Pinus taeda* en donde se calculó que la especie portaba al menos ocho equivalentes letales (es decir, aquellos alelos que pueden causar la muerte en forma homóciga) y una depresión del 98% en la progenie endógama (Remington y O'Malley, 2000). Posiblemente hasta la fecha se haya subestimado el impacto de la DE porque en muchas ocasiones los genes letales y sub-letales se expresan en estadíos tempranos del desarrollo y sólo se desarrollan los heterócigos, que son los que nosotros muestreamos, dando bajos niveles de endogamia. Keller y Waller (2002), demuestran que hay DE en alrededor de 40 especies de animales y plantas.

Las consecuencias de la DE abarcan diferentes aspectos de la biología de los individuos como reducción en la fecundidad y la sobrevivencia. Se ha observado que en aves y mamíferos la DE afecta el peso al nacer, la sobrevivencia, la reproducción y la resistencia a enfermedades, así como al estrés ambiental y a la depredación (Crnokrak y Roff, 1999; Hedrick y Kalinowski, 2000). En plantas, se ve afectada la producción de semillas viables, la germinación, la sobrevivencia y la resistencia al estrés (Keller y Waller, 2002).

Existen dos ejemplos interesantes en los que se ha demostrado la depresión por endogamia en ambientes naturales. La planta *Clarkia pulchella* demostró un mayor riesgo de extinción debido a la disminución de los tamaños efectivos y al aumento de la endogamia (Newman y Pilson, 1997). El otro trabajo fue sobre una metapoblación de mariposas del género *Melitaea* (Saccheri *et al.*, 1998), donde se mostró que el riesgo de extinción aumentaba considerablemente si disminuía la heterocigosis y aumentaba la endogamia. En este mismo trabajo se encontró que los componentes de la adecuación que se modificaban con la baja heterocigosis eran la sobrevivencia de las larvas, la longevidad de los adultos y el número de huevos que eclosionan.

Con simulaciones en computadora se ha demostrado que la DE es un factor probable que aumenta el riesgo de extinción en poblaciones que normalmente son exógamas en la naturaleza (Mills y Smouse, 1994).

También se ha demostrado que la endogamia hace a los individuos más susceptibles a la mortalidad debida al ambiente (Reed *et al.*, 2002). Hay dos ejemplos interesantes al respecto; en borregos de la especie *Ovis aries* los homócigos eran más susceptibles al parasitismo y su sobrevivencia menor (Coltman *et al.*, 1999). Otro caso es el de los gorriones *Melospiza melodia,* donde las aves endógamas morían con mayor frecuencia durante tormentas

severas. Aunque la muerte era causada por las tormentas, la endogamia determinaba quiénes eran los sobrevivientes (Keller *et al.*, 1994).

¿Y si se purgan las poblaciones?

En los experimentos en los que por primera vez Darwin notó la depresión por endogamia realizando cruzas controladas en *Ipomea purpurea*, notó que después de varias generaciones de autocruza, surgían plantas con un sorprendente vigor a los que él llamó plantas "héroe". Los descendientes de estos héroes mantenían este vigor y Darwin concluyó que los cambios deletéreos, que inicialmente producían depresión por endogamía eran eliminados de alguna manera, siendo éste justamente el efecto de la purga (Byers y Waller, 1999). Actualmente sabemos que este desarrollo superior de algunos individuos resultado de autocruzas como el héroe de Darwin y la reducción de la depresión por endogamia en líneas autógamas son el resultado de la eliminación preferencial de alelos recesivos letales (Byers y Waller, 1999). La endogamia expone rápidamente los recesivos letales en los individuos homócigos y la selección natural reduce su frecuencia. Las mutaciones deletéreas en recesivos se purgan más fácilmente en poblaciones pequeñas o endógamas que en poblaciones grandes con apareamiento al azar (Charlesworth y Charlesworth, 1999).

Una idea que se adoptó en los programas de manejo de especies en peligro fue forzar la endogamia para purgar las mutaciones deletéreas de la población. Aunque en el caso en que la depresión por endogamia sea resultado de muchos alelos deletéreos con efecto menor, el forzar la endogamia provocará que se fijen estos alelos en lugar de purgarlos, produciendo una disminución en la adecuación global (Hedrick y Kalinowsky, 2000).

A la fecha se han hecho varias revisiones importantes del efecto de la purga, Ballou (1997) examinó 25 estudios sobre mamíferos y Byers y Waller (1999) 52 sobre plantas, y concluyeron que la purga sólo reduce la depresión por endogamia en algunos caracteres, en algunas poblaciones y en un grado limitado.

Hibridación e introgresión

La introgresión es la introducción en una población de material genético de otras especies o subespecies; dando como resultado una hibridación entre dos formas previamente diferenciadas (véase el capítulo 13 de este libro). Este proceso puede amenazar la integridad genética de una población, cuando son in-

troducidas especies exóticas dentro del rango de distribución de especies raras o cuando hay una alteración del hábitat en donde especies que se encontraban aisladas entran en contacto e hibridizan. Una forma particular de introgresión deletérea, en el caso de plantas, ocurre cuando entran en contacto poblaciones tetraploides y diploides, dando como resultado poblaciones triploides estériles (Young y Murray, 2000). También se ha observado la presencia de genes de perros en varias especies de lobos (Randi y Lucchini, 2002). Para remediar los efectos causados por la introgresión es necesario en principio eliminar la especie introducida o translocar individuos puros a regiones aisladas o de cautiverio, para después volver a fundar la población.

Otro ejemplo es el de *Gliricidia sepium,* una leguminosa que se distribuye desde México hasta Centroamérica, con gran importancia económica porque mejora la productividad de los sistemas agrícolas. Dado que se entrecruza con la especie *G. maculata*, que se distribuye en la península de Yucatán, Dawson y colaboradores (1996) investigaron las posibles interacciones entre ambas especies para poder definir estrategias de conservación y manejo. Ambas especies presentan plasticidad morfológica (lo que hace difícil identificar híbridos), por lo que las técnicas moleculares permitieron determinar el flujo génico dentro y entre las poblaciones así como la posible introgresión. Se identificó una zona híbrida que concuerda con la dispersión antropogénica, lo cual representaría un mecanismo de erosión genética que en un futuro puede poner en riesgo a las especies.

Los enfoques actuales de la genética de la conservación

Con los nuevos avances en los métodos de análisis genéticos y la gran cantidad de información que se ha obtenido con la liberación de muchas secuencias de ADN e genomas completos, no sólo se ha revolucionado la genética de poblaciones, sino su aportación a la conservación. Además, cada vez parece más claro que en los programas de genética de la conservación deben ser incorporados datos de la taxonomía y ecología de las especies, además del empleo de nuevas herramientas moleculares y computacionales para determinar el estatus de las especies que se quieren conservar.

Desde el punto de vista taxonómico se ha enfatizado en conservar aquellos taxa que contribuyen a la biodiversidad global en una escala proporcional a sus características distintivas tanto morfológicas como genéticas (por ejemplo organismos como el celacanto o el tuátara). Con el desarrollo de métodos

filogenéticos se pueden hacer comparaciones de secuencias y calcular el largo de las ramas en los árboles filogenéticos para estimar qué tan distintivos son los taxa (Bowen, 1999). Por otro lado, los conocimientos actuales de genomas y el desarrollo de la teoría de evolución molecular han permitido aumentar el conocimiento sobre la diversidad genética, así como como identificar a los grupos con alta capacidad de producir nuevas especies que serán la fuente de futura biodiversidad (Erwin, 1991).

Otro enfoque que ha tomado auge en los últimos años es el de conservar los procesos evolutivos y mantener la capacidad de respuesta evolutiva ante los cambios ambientales (Erwin, 1991). Por ejemplo, Moritz (2002) propone que generalmente existen zonas donde está bien representada la diversidad genética, y en estas áreas se debe maximizar la protección de gradientes ambientales continuos, en donde interactúan la selección y la migración para mantener la viabilidad de las poblaciones.

Unidades Evolutivamente Significativas y Unidades de Manejo

En años recientes el problema de definir las prioridades de conservación ha recaído un poco más en el área de la sistemática para tratar de entender las "unidades evolutivas" que deben ser conservadas. Durante mucho tiempo en biología ha existido un debate sobre qué es una especie y sobre los diversos conceptos de especie que se han generado (véase el capítulo 10 de este libro). La mayoría de especies prioritarias se han propuesto a partir de un concepto tipológico, lo que trae como desventaja el que no se considere la diversidad real dentro de las especies, ni la variación geográfica, ni la posibilidad de mantener su potencial evolutivo (Rojas, 1992).

En el contexto de la conservación se requiere delimitar ciertas unidades de conservación, que han sido definidas como *unidades evolutivamente significativas* (o *ESU*, por sus siglas en inglés) y que son "unidades poblacionales que merecen manejo propio y que tienen una alta prioridad de conservación" según la definición de Ryder (1986). Estas unidades deben darnos una idea acerca de los procesos evolutivos y la distribución de la diversidad genética para llevar a cabo estrategias de conservación. Las ESUs no tienen que coincidir forzosamente con la categoría de especie, ya que pueden tratarse de poblaciones distintas genéticamente del resto o que han estado históricamente aisladas (Soltis y Gitzendanner, 1999; Fraser y Bernatchez, 2001). En este sentido, la diferenciación dentro de las especies hace que se tengan que conservar más

poblaciones para asegurar que se mantenga suficiente variación genética para asegurar la superviviencia de la especie.

La categoría de *unidad de manejo* (MU), por otra parte, ha servido para reconocer poblaciones demográficamente distintas que deben ser manejadas para asegurar la viabilidad de las ESU que son más grandes. Las unidades de manejo se han utilizado principalmente cuando se conoce la estructura poblacional actual y no la estructura histórica (Moritz, 1994).

Por ejemplo, la serpiente mocasín mexicana *Agkistrodon bilineatus* presenta una considerable variación en el patrón de color, lo que representa un problema taxonómico. Parkinson y colaboradores (2000) mostraron con datos moleculares que la subespecie *Agkistrodon bilineatus taylori* merecía ser considerada como una unidad aparte (y no una subespecie), lo que hace urgente un plan de conservación para ella, ya que además se encuentra en grave amenaza (mucho más que otras variedades). Se ha observado que la mocasín mexicana puede ser criada con éxito en cautiverio, aunque se debe poner énfasis en mantener a las distintas variedades como unidades separadas.

Por otro lado, la musaraña *Sorex ornatus* es una especie rara que está subdividida en nueve subespecies (Owen y Hoffman, 1983), de las cuales siete tienen distribuciones en pequeños parches, *S. o. ornatus* se encuentra en Baja California Norte y *S. o. lagunae* habita en la punta de Baja California Sur. Su distribución presenta barreras geográficas que generan diferencias genéticas, además de que sus hábitos muestran altos grados de especialización, por lo que las poblaciones se encuentran muy aisladas y con dispersión muy limitada. Con técnicas moleculares se demostró que las poblaciones son genéticamente divergentes, aunque se encuentren geográficamente cerca, y se encontró que existen tres poblaciones que forman clados bien definidos (Maldonado *et al.*, 2001). Se propone que la especie tuvo alguna vez una distribución continua y que la explotación de las tierras y mantos acuíferos fue aislando a las poblaciones, al grado que las poblaciones actuales de Baja California son muy pequeñas y fragmentadas y tienen una alta probabilidad de extinción. Por ello se sugiere que los clados encontrados deben ser manejados como unidades independientes, ya que pueden responder de manera independiente a los cambios ambientales, es decir, que si una población se extingue, difícilmente va a poder recolonizarse la zona donde habitaba. Hay que considerar que las poblaciones mexicanas, además de ser las más amenazadas también son las más distintas genéticamente, por lo que es imperativo realizar acciones que permitan mantener sus hábitats para proteger a las pocas poblaciones existentes. Adicionalmente se encontró que los resultados no concuerdan con los

estatus taxonómicos propuestos, por lo que es necesaria una revisión para determinar con mayor certeza las unidades de manejo.

Para el caso del borrego cimarrón *Ovis canadensis* se han hecho muchos esfuerzos por mantener tanto la especie como sus subespecies (tradicionalmente siete), sin embargo el valor taxonómico de las subespecies es bastante dudoso, lo cual resulta muy importante para los programas de reintroducción. Ramey (1995) encontró que no hay datos moleculares que apoyen la existencia de cinco de la siete subespecies, lo cual puede simplificar los programas de reintroducción.

Metapoblaciones

La estructura poblacional ha tenido una repercusión muy importante en los modelos de biología y genética de poblaciones. *Metapoblación* puede referirse casi a cualquier población estructurada espacialmente, aunque la definición tradicional es la de Levins (1969): conjunto de poblaciones inestables, que ocupan parches o segmentos de hábitat discretos, que varían en área, grado de aislamiento y calidad y cantidad de los recursos que requiere dicha especie. En otras palabras, las metapoblaciones son un mosaico cambiante de poblaciones temporales interconectadas por algún grado de migración (Levins, 1969). Una metapoblación clásica persiste por un balance entre las "muertes " (extinciones locales) y los "nacimientos" (establecimiento de nuevas poblaciones en sitios desocupados previamente) (Hanski, 1998). Dado que en las metapoblaciones los eventos de extinción son recurrentes, y no eventos únicos y aislados, nos permiten comprender mejor el fenómeno de la extinción, ya que dependen, al igual que en las poblaciones pequeñas, de la estocasticidad demográfica, genética y ambiental. El conocimiento acerca de los eventos de colonización y migración permite considerar el **efecto fundador** y de deriva. Existen ejemplos como el mono tití del Amazonas (*Callicebus bernhardi*), en donde el enfoque metapoblacional ha permitido mantener algunas poblaciones muy pequeñas en fragmentos ya deteriorados del Amazonas y por otro lado una población núcleo en cautiverio (Primack *et al.*, 2001).

La fragmentación del hábitat puede crear una metapoblación de lo que fue un gran población con distribución continua y sin pasar por cuellos de botella se puede perder una gran diversidad (Gilpin, 1991; Hanski y Gilpin, 1997).

Filogeografía

Con el desarrollo de técnicas moleculares que estudian el ADN de mitocondria y de cloroplasto, se han construido filogenias de haplotipos en las que se examina la distribución geográfica de la variación genética; a esto se le conoce como filogeografía molecular (véase los capítulos 14 y 15 de este libro). Dado que la aproximación de la filogeografía se basa más en la representatividad de los linajes, pueden protegerse los principales linajes históricos dentro de las especies, lo que tiene gran importancia porque una vez perdidos los linajes ya no pueden ser recuperados (Avise, 2000). La filogeografía también ha permitido, en estudios de animales, dilucidar cómo la historia evolutiva reciente ha moldeado los patrones de diversidad dentro de las especies (Avise, 1998), por ejemplo debido a eventos como las expansiones del rango de distribución, la fragmentación y los cuellos de botella, que parecen haber tenido una mayor influencia en los patrones de diferenciación genética que lo que se había calculado previamente con modelos tradicionales de genética de poblaciones (Schaal *et al.*, 1998).

La filogeografía ha mejorado mucho la descripción de la distribución geográfica, las relaciones filogenéticas y distancias genéticas entre linajes de animales, logrando tener una mejor comprensión de la biogeografía regional y áreas de endemismo. Justamente, comparar los patrones filogeográficos de varios grupos en una misma región es cuando podemos comprender la historia de dicha región, lo que se conoce como biogeografía comparada (Bermingham y Moritz, 1998).

Por otro lado, comprender las respuestas históricas de las especies a cambios en el ambiente y la evolución de áreas evolutivamente aisladas es de gran importancia para la conservación, ya que puede ayudar a planear estrategias de conservación (Bermingham y Moritz, 1998; Moritz y Faith, 1998). Por ejemplo, la distribución original del jaguar (*Panthera onca*) abarcaba desde el sur de Estados Unidos hasta la Patagonia, pero ha disminuido recientemente y en la actualidad existen poblaciones fragmentadas de tamaño variable, que se distribuyen desde México hasta Brasil. Eizirik y colaboradores (2001) presentaron un estudio filogeográfico sobre su historia poblacional y diversidad genética, encontrando que la especie es joven y que ha tenido una reciente expansión poblacional, pero que el tiempo ha sido insuficiente para generar diferenciación regional, lo que a su vez sugiere que mantuvieron altos niveles de flujo genético en grandes distancias. No se detectó evidencia alguna que sugiriera la reducción de variación genética, aunque esto puede estar influido por el reciente origen de la especie; además,

no encontraron subdivisiones que indiquen subespecies, lo que se contrapone con la clasificación de algunos autores. Así, este trabajo sugiere que los planes de manejo deben permitir la existencia de flujo génico entre poblaciones que se encuentren aisladas, así como entre ejemplares que se encuentren en cautiverio, aunque también se deben considerar posibles adaptaciones locales que no hayan sido detectadas por los marcadores moleculares.

Conclusiones y perspectivas

Pese al debate que ha existido entre la importancia de los procesos genéticos en el campo de la biología de la conservación, es claro ahora que la demografía, ecología y genética de poblaciones interactúan y afectan la persistencia de las poblaciones pequeñas (Eguiarte y Piñero, 1990). Hoy en día existe evidencia de que los factores genéticos están involucrados en el riesgo de extinción en las especies, aunque el efecto ocurre a diferentes escalas temporales. Por ejemplo, el efecto más rápido lo podría tener la deriva génica, ya que al aumentar las cruzas entre parientes y la fijación de la carga genética, se reduce la adecuación y se produce la depresión por endogamia (Keller y Waller, 2002).

Tal vez el papel de la endogamia se ha subestimado al no conocer exactamente su interacción con la selección y su efecto en la dinámica poblacional. Los últimos trabajos sobre depresión por endogamia dejan abierto un campo de estudio muy amplio porque aún no se han encontrado patrones generales de cómo la DE varía entre diferentes taxa, ambientes y poblaciones con historias genéticas y demográficas contrastantes (Crnokrak y Roff, 1999; Hedrick y Kalinowski, 2000; Keller y Waller, 2002).

Otra recomendación importante es conocer los niveles de diversidad y lo que éstos pueden decirnos sobre el riesgo de especies amenazadas; un ejemplo claro es la falta de información sobre caracteres cuantitativos, en particular en las especies en peligro que han mostrado bajos niveles de variación. La baja correlación entre caracteres moleculares y cuantitativos da la idea de que los primeros no pueden ser usados como el único medio para evaluar la diversidad, ya que no necesariamente reflejan el potencial evolutivo de las poblaciones, principalmente para los caracteres relacionados con la historia de vida o los caracteres adaptativos. Las especies en peligro han mostrado niveles de variación bajos, pero son muy pocos los estudios que han medido la variación cuantitativa. Sería importante poder comparar desde esta perspectiva la reducción en el potencial evolutivo entre especies en peligro y sus congéneres no en peligro (Reed y Frankham, 2001).

También es necesario comprender de mejor manera cómo los diferentes procesos genéticos llevan al llamado vórtice de extinción.

Por otro lado, los principios de la genética de la conservación deben ser evaluados de manera especial para cada especie, ya que se ha visto que el efecto de la reducción en los tamaños poblacionales depende del sistema reproductivo, la selección natural, la demografía y la historia de las especies (Moritz *et al.*, 1993).

Por su parte, los enfoques nuevos como las herramientas filogeográficas, o la conservación de la complejidad taxonómica (Ennos *et al.*, 2005), nos permiten evaluar de manera más integral los diferentes procesos evolutivos, para que la genética de la conservación pueda brindar las herramientas necesarias a los manejadores y aportar información práctica relevante para priorizar y medir la biodiversidad (Moritz, 2002).

Bibliografía

Amos W. y A. Balmford. 2001. When does conservation genetics matter? *Heredity* 87:257-265.

Amos W. y J. Harwood. 1998. Factors affecting levels of genetic diversity in natural populations. *Phil. Trans. Royal Soc. London B* 353:177-186.

Avise J. C. 1998. The history and purview of phylogeography: a personal reflection. *Mol. Ecol.* 7:371-379.

Avise J. C. 2000. *Phylogeography, the history and formation of species*. Harvard University Press, Cambridge.

Baker C.S., G.M. Lento, F. Cipriano y S.R. Palumbi. 2000. Predicted decline of protected whales based on molecular genetic monitoring of Japanese and Korean markets. *Proc. R. Soc. Lond. Ser. B-Biol. Sci.* 267:1191-1199.

Ballou J. 1997. Ancestral inbreeding only minimally affects inbreeding depression in mammalian populations. *J. Hered.* 88: 169-178.

Baur B. y B. Schmid. 1996. Spatial and temporal patterns of genetic diversity within species. En: Gaston, K. J. (ed). Biodiversity: a biology of numbers and difference. Blackwell Science, Reino Unido.

Baur B. y M. Klemm. 1989. Absence of isozyme variation in geographically isolated population of the land snail *Chondrina clienta*. *Heredity* 63: 239-244.

Beaumont M.A. 2001. Conservation genetics. En: Balding *et al.*, (eds.). *Handbook of statistical genetics.* John Wiley and Sons. pp. 779-811.

Bermingham E. y C. Moritz. 1998. Comparative phylogeography: concepts and applications. *Mol. Ecol.* 7:367-369.

Bevill R.L. y S.M. Louda. 1999. Comparisons of related rare and common species in the study of plant rarity. *Cons. Biol.* 13:493-498.

Bowen B.W. 1999. Preserving genes, species, or ecosystems: Healing the fractured foundations of conservation policy. *Mol. Ecol.* 8:S5-S10.

Byers D.L. y D.M. Waller. 1999. Do plant populations purge their genetic load? Effects of population size and mating history on inbreeding depression. *Annu. Rev. Ecol. Syst.* 30:479-513.

Caro T.M. y M.K. Laurenson. 1994. Ecological and genetic factors in conservation: a cautionary tale. *Science* 263:485-486.

Caughley G. 1994. Directions in conservation biology. *Journal of Animal Ecology* 63:215-244.

Charlesworth B. y D. Charlesworth. 1999. The genetic basis of inbreeding depression. *Genet. Res. Camb.* 74:329-340.

Coltman D.W., J.G. Pilkington, J. A. Smith y J. M. Pemberton. 1999. Parasite-mediated selection against inbred Soay sheep in a free-living, island population. *Evolution* 53:1259-1267.

Crnokrak P. y D.A. Roff. 1999. Inbreeding depression in the wild. *Heredity* 83:260-270.

Darwin C. 1876. *The effects of cross and self fertilization in the vegetable kingdom*. J. Murray. Londres.

Dawson I., A. Simons, R. Waugh y W. Powel. 1996. Detection and pattern of interes-pecific hybridization between *Gliciridia sepium* and *G. maculata* in Meso-America revealed by PCR-based analysis. *Mol. Ecol.* 5:89-98.

Dinerstein, E. y G. F. McCracken. 1990. Endangered greater one-horned rhinoceros carry high levels of genetic variation. *Cons. Biol.* 4: 417-422.

Eguiarte, L. E. y D. Piñero. 1990. Génetica de la conservación: leones vemos, genes no sabemos. *Ciencias*. Número especial 4. Ecología y conservación en México: 34 –47 (reimpreso en Nuñez-Farfán J. y L. E. Eguiarte (editores). *La evolución biológica*. Facultad de Ciencias, Instituto de Ecología, UNAM, Conabio. Pp. 371-398).

Eizirik E., J.H. Kim, M. Menotti-Raymond, P.G. Crawshaw, S.J. O'Brien y W.E. Johnson. 2001. Phylogeography, population history and conservation genetics of jaguars (*Panthera onca*, Mammalia, Felidae). *Mol. Ecol.* 10:65-79.

Ennos, R. A., G. C. French y P. M. Hollingsworth. 2005. Conserving taxonomic complexity. *Trends. Ecol. Evol.* 20: 164-167.

Erwin T. L. 1991. An evolutionary basis for conservation strategies. *Science* 253:750-752.

Falconer, D. S. y T. F. C. Mackay. 1996. *Introduction to quantitative genetics*. 4th Edition. Addison Wesley Longman, New York.

Falk, D. A. y K. E. Holsinger. 1991. *Genetics and conservation of rare plants*. Oxford University Press, New York.

Frankham R. 1995. Conservation genetics. *Annu. Rev. Genet.* 29:305-327.

Frankham R., J.D. Ballou, y D.A. Briscoe. 2002. *Introduction to conservation genetics*. Cambridge, Reino Unido.

Franklin I.R. 1980. Evolutionary change in small populations. En: Soulé M. E. y B. A. Wilcox (eds.). *Conservation Biology*. Sinauer Associates, Sunderland, Mass. pp. 135–150.

Fraser D.J. y L. Bernatchez. 2001. Adaptive evolutionary conservation: towards a unified concept for defining conservation units. *Mol. Ecol.* 10:2741-2752.

Gilpin M. 1991. The genetic effective size of a metapopulation. *Biol. J. Linn. Soc.* 42:165-175.

Gitzendanner M. A. y P. Soltis. 2000. Patterns of genetic variation in rare and widespread plant congeners. American Journal of Botany 87(6):783-792.

Hamrick J.L. y M.J.W. Godt. 1990. Allozyme diversity in plant species. En: A.H.D. Brown, M.T. Clegg, A.L. Kahler, y B.S. Weir. (eds.). *Plant population genetics, breeding, and genetic resources*. Sinauer, Massachusetts. pp. 43–63.

Hanski I. 1998. Metapopulation dynamics. *Nature* 396:41-49.

Hanski I. y M. Gilpin. 1997. *Metapopulation biology*. Academic Press, San Diego.

Hartl, D.L. y A.G. Clark. 1997. *Principles of population genetics*. Sinauer, Sunderland, Mass.

Hedrick P.W. 1995. Gene flow and genetic restoration: the Florida panther as a case study. *Cons. Biol.* 9:996-1007.

Hedrick P.W. 2000. *Genetics of populations*. Jones and Bartlett Pub, Mass.

Hedrick P.W. 2001. Conservation genetics: where are we now? *Trends. Ecol. Evol.* 16:629-636.

Hedrick P.W. y S. T. Kalinowski. 2000. Inbreeding depression in conservation biology. *Annu. Rev. Ecol. Syst.* 31:139-162.

Hedrick, P. W. 2005. 'Genetic restoration': a more comprehensive perspective than 'genetic rescue' *Trends. Ecol. Evol.* 20: 109.

Hoelzel A.R., J. Halley, S.J. O'Brien, C. Campagna, T. Arnborm, B. Le Boeuf, K. Ralls y G.A. Dover. 1993. Elephant seal genetic variation and the use of simulation models to investigate historical population bottlenecks. *J. Heredity* 84:443.

Husband B.C. y D.W. Schemske. 1996. Evolution of the magnitude and timing of inbreeding depression in plants. *Evolution* 50:54-70.

Ingvarsson P.K. 2001. Restoration of genetic variation lost –the genetic rescue hypothesis. *Trends. Ecol. Evol.* 16:62-63.

Karron J.D. 1987. A comparison of levels of genetic polymorphism and self-compatibility in geographically restricted and widespread plant congeners. *Evolutionary Ecology* 1:47–58.

Keller L.F. y D.M. Waller. 2002. Inbreeding effects in wild populations. *Trends. Ecol. Evol.* 17:230-241.

Keller L.F., P. Arcese, J.N.M. Smith, W.M. Hochachka y S.C. Stearns.1994. Selection against inbred song sparrows during a natural population bottleneck. *Nature* 372:356-357.

Kunin W.E. y K.J. Gaston. 1993. *The biology of rarity: causes and consequences of rare-common differences*. Chapman and Hall, Londres.

Lande R. 1988.Genetics and demography in biological conservation. *Science* 241:1455-1460.

Lande R. 1995. Mutation and conservation. *Conservation Biology* 9:82-791.

Lande R. y G.F. Barrowclough. 1987. Effective population size, genetic variation, and their use in population management. En: Soule M. E. (ed.). *Viable populations for conservation*. Cambridge University Press, Cambridge. pp. 87-123.

Larson, S., R. Jameson, M. Etnier, M. Fleming y P. Bentzen. 2002. Loss of genetic diversity in sea otters (*Enhydra lutris*) associated with the fur trade in the 18th and 19th centuries. *Mol. Ecol.* 11: 1899-1903.

Levins R. 1969. Some demographic and genetic consequences of environmental heterogeneity for biological control. *Bull. Entomol. Soc. Am.* 15:237-240.

Loveless M.D. y J.L. Hamrick. 1984. Ecological determinants of genetic structure in plant populations. *Annu. Rev. Ecol. Syst.* 15:65-95.

Lynch M. 1993. The mutational meltdown in asexual populations. *J. Hered.* 84:339-344.

Lynch M., J. Conery y R. Bürger. 1995. Mutational meltdowns in sexual populations. *Evolution* 49:1067-1080.

Maldonado J.E., C. Vilà y R.K. Wayne. 2001. Tripartite genetic subdivision in ornate shrew (*Sorex ornatus*). *Mol. Ecol.* 10:127-147.

McNeely J.A., K.R.Miller, W.V. Reid, R.A. Mittermeier y T.B. Werner. 1990. *Conserving the world's biological diversity.* IUCN, World Resources Institute, Conservation International, WWF-US, World Bank, Washington.

Meffe G.K. y R. Carroll. 1994. *Principles of conservation biology*. Sinauer, Mass.

Mills L.S. y P.E. Smouse. 1994. Demographic consequences of inbreeding in small remnant populations. *Am. Natur.* 144:412-431.

Moritz, C. 1994. Defining "Evolutionary significant units" for conservation. *Trends Ecol. Evol.* 9: 373-375.

Moritz C. 2002. Strategies to protect biological diversity and the evolutionary processes that sustain it. *Syst. Biol.* 51(2):238-254.

Moritz C. y D.P. Faith. 1998. Comparative phylogeography and the identification of genetically divergent areas for conservation. *Mol. Ecol.* 7:419-429.

Moritz C., D.J. Coates, W. Sherwin, T. Clancy y C.J. Limpus. 1993. Population ecology and genetics. En: Moritz, C y J. Kikkawa (eds.). *Conservation Biology in Australia and Oceania*. Surrey Beatty & Sons, Australia. pp.359-362.

Nevo E. 1983. Adaptive significance of protein variation. En: Oxford, G.S. y D. Rollinson (eds.). *Protein polymorphism: adaptive and taxonomic significance*. Academic Press, Londres. pp. 240–282.

Newman D. y D. Pilson. 1997. Increased probability of extinction due to decreased genetic effective size: experimental populations of *Clarkia pulchella*. *Evolution* 51:354-362.

O'Brien S. J. 1994. A role for molecular genetics in biological conservation. *Proc. Natl. Acad. Sci.* 91:5748-5755.

Owen, J. G., y R. S. Hoffmann. 1983. *Sorex ornatus*. *Mammalian Species* 212: 1-5.

Parkinson C.L., K. Zamudio y H.H. Greene. 2000. Phylogeography of the pitviper clade *Agkistrodon*: historical ecology, species status, and conservation of cantils. *Mol. Ecol.* 9:411-420.

Primack R. 2001. Problemas de las poblaciones pequeñas. En: R. Primack *et al.* (eds.). *Fundamentos de conservación biológica*. Fondo de Cultura Económica, México. pp. 363-383.

Primack R., R. Rozzi, P. Feisinger, R. Dirzo y F. Masardo. 2001. *Fundamentos de conservación biológica*. Fondo de cultura económica, México.

Rabinowitz D. 1981. Seven forms of rarity. En: H. Synge (ed.). *The biological aspects of rare plant conservation*. Wiley, Nueva York. pp.205-217.

Randi E. y V. Lucchini. 2002. Detecting rare introgression of domestic dog genes into wild wolf (*Canis lupus*) populations by Bayesian admixture analyses of microsatellite variation. *Cons. Gen.* 3:31- 45.

Ramey II R. 1995. Mitochondrial DNA variation, population structure, and evolution of mountain sheep in the south western United States and Mexico. *Mol. Ecol.* 4:429-439.

Reed D.H. y R. Frankham. 2001. How closely correlated are molecular and quantitative measures of genetic variation? A meta-analysis. *Evolution* 55:1095-1103.

Reed D.H. y R. Frankham. 2003. Correlation between fitness and genetic diversity. *Cons. Biol.* 17:230-237.

Reed D.H., D.A. Briscoe y R. Frankham. 2002. Inbreeding and extinction: The effect of environmental stress and lineage. *Cons. Gen.* 3:301-307.

Remington D.L. y D.O.O'Malley. 2000. Whole-genome characterization of embryonic stage inbreeding depression in a selfed loblolli pine family. *Genetics* 155:337-348.

Rojas M. 1992. The species problem and conservation: what are we protecting? *Cons. Biol.* 6:170-178.

Ryder O.A. 1986. Species conservation and systematics: the dilemma of subspecies. *Trends. Ecol. Evol.* 1:9-10.

Saccheri I., M. Kuussaari, M. Kankare, P. Vikman, W. Fortelius y I. Hanski. 1998. Inbreeding and extinction in a butterfly metapopulation. *Nature* 392:491-494.

Sawkins M., B.L. Maass, B.C. Pengelly, H.J. Newbury, B.V. Ford-Lloyd, N. Maxted y R. Smith. 2001. Geographical patterns of genetic variation in two species of *Stylosanthes* Sw. using amplified fragment length polymorphism. *Mol. Ecol.* 10:1947-1958.

Schaal B.A., D.A. Hayworth, K.M. Olsen, J.T. Rauscher y W.A. Smith. 1998. Phylogeographic studies in plants: problems and prospects. *Mol. Ecol.* 7:465-474.

Shivji M., S. Clark, M. Pank, L. Natanson, N. Koheler y M. Stanhope. 2002. Genetic identification of pelagic shark body parts for conservation and trade monitoring. *Cons. Biol.* 16:1036-1047

Simberloff D. 1988. The contribution of population and community biology to conservation science. *Ann. Rev. Ecol. Syst.* 19:473-511.

Soltis P.S. y M.A. Gitzendanner. 1999. Molecular systematics and the conservation of rare species. *Cons. Biol.* 13:471–483.

Soulé M.E., M. Gilpin, W. Conway y T. Foose. 1986. The millenium ark: how long a voyage, how many staterooms, how many passengers? *Zool. Biol.* 5:101-113.

Spielman D. y R. Frankham. 1992. Modeling problemas in conservation genetics using captive *Drosophila* populations: improvement of reproductive fitness due to immigration of one individual into small partial inbred populations. *Zool. Biol.* 11:343-351.

Storfer A. 1996. Quantitative genetics: a promising approach for the assessment of genetic variation in endangered species. *Trends. Ecol. Evol.* 11:343-348.

UICN. 2001. *Categorías y criterios de la lista roja de la UICN: Versión 3.1.* Comisión de Supervivencia de Especies de la UICN. UICN, Gland, Suiza y Cambridge, Reino Unido. 33 pp.

Vilà C., A.K. Sundqvist, O. Flagstad, J. Seddon, S. Bjornerfeldt, I. Kojola, A. Casulli, H. Sand, P. Wabakken y H. Ellegren. 2003. Rescue of a severely bottlenecked wolf (*Canis lupus*) population by a single immigrant. *Proc. R. Soc. Lond. Ser. B-Biol. Sci.* 270:91-97.

Young A.G. y A.H.D. Brown. 1996. Comparative population genetic structure of the rare woodly shrub *Daviesia suaveolens* and its common congener *D. mimosoides*. *Cons. Biol.* 10:1220–1228.

Young A.G. y B.G. Murray. 2000. Genetic bottlenecks and dysgenic gene flow into re-established populations of the grassland daisy *Rutidosis leptorrynchoides*. *Aust. J. Bot.* 48:409-416.

Westemeier R.L., J.D. Brawn, S.A. Simpson, T.L. Esker, R.W. Jansen, J.W. Walk, E.L. Kershner, J.L. Bouzat y K.N. Paige. 1998. Tracking the long-term decline and recovery of an isolated population. *Science* 282:1695-1698.

Tercera parte

La ecología molecular de los microorganismos

Capítulo 9

La recombinación: relevancia evolutiva y métodos de estimación, con énfasis en microorganismos

Rodolfo Salas Lizana

La recombinación es, en un sentido amplio, el intercambio de material genético entre genomas diferentes. Es un fenómeno común en todos los organismos, incluyendo los virus (Holmes *et al.*, 1999); también se han reportado casos en los organelos mitocondria (Saville *et al.*, 1998) y cloroplasto (Marshall *et al.*, 2001).

La recombinación modifica profundamente los genomas. Su principal contribución en los eucariontes es la de generar combinaciones de alelos en nuevos genomas (p.ej. Zhan *et al.*, 2003). Además, principalmente en los procariontes, permite integrar nuevos genes en los organismos (Kooning *et al.*, 2001; Ochman *et al.*, 2000), así como eliminar a otros (Ham *et al.*, 2003; Lawrence *et al.*, 2001), regulando el tamaño del genoma. Es por esto que la recombinación es invocada cada vez más para explicar los patrones de variación observados en muchos genomas (Akhunov *et al.*, 2003; Boucher *et al.*, 2001; Edwards y Brookfield, 2003; Heidelberg *et al.*, 2000; Parkhill *et al.*, 2001; Raymond *et al.*, 2002; Suyama y Bork, 2001).

Por otra parte, estimar o detectar la presencia de la recombinación en secuencias de ADN es de mucha utilidad, de tal manera que, como otros parámetros de genética de poblaciones, nos puede ayudar a inferir qué tanto se intercambian los genes entre los genomas en una población, entre poblaciones de una misma especie y entre especies diferentes (Awadalla, 2003), lo que deriva directamente en un mejor entendimiento de las dinámicas microbianas,

sobre todo de las que influyen directamente en la cotidianeidad humana. Asimismo, en un sentido más práctico, es una herramienta que permite mapear genes de fenotipos de interés, como la virulencia, resistencia a enfermedades o males genéticos en humanos.

Mecanismos de la recombinación en eucariontes

En la meiosis, la recombinación puede ocurrir por rearreglos independientes de genes en diferentes cromosomas (descritos por la segunda ley de Mendel; figura 1) o por entrecruzamiento (*crossing-over* en inglés) entre genes homólogos en cromátidas no hermanas (Griffiths *et al.*, 1999).

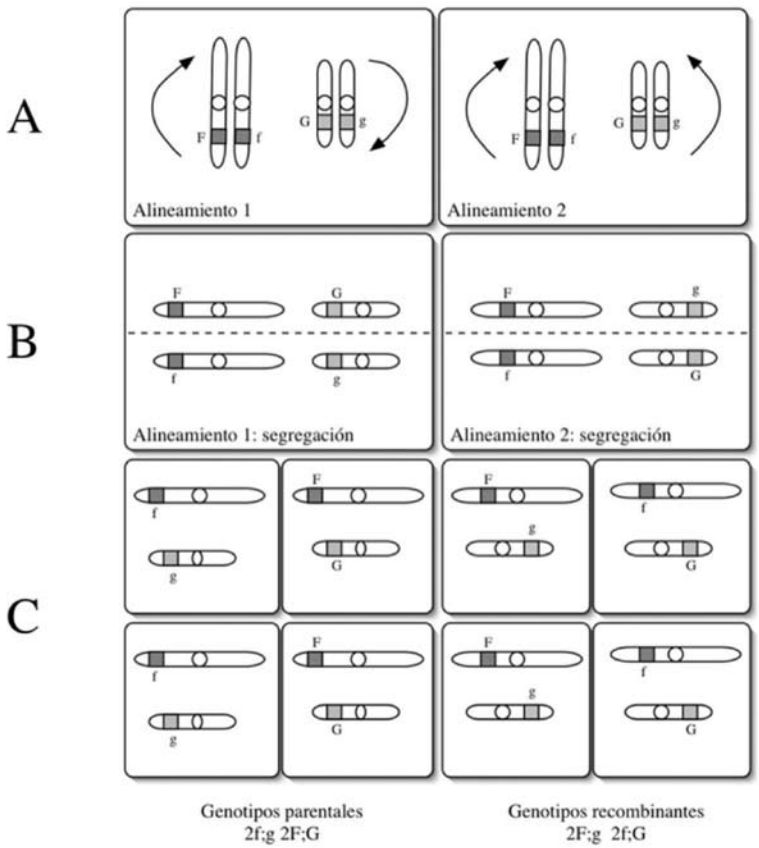
El arreglo independiente siempre produce el 50% de genotipos recombinantes, mientras que el porcentaje puede variar en el *crossing-over* entre 0 y 50%. Esto es porque durante un evento meiótico es posible que dos genes homólogos no se entrecrucen o que intercambien fragmentos de ADN de tamaño variable (figura 2).

El arreglo independiente puede producir genomas recombinantes al ocurrir ordenamientos alternativos de las cromátidas hermanas antes de la haploidación en la mitosis II (figura 1), sin que ocurra entrecruzamiento de ADN entre cromosomas.

Durante la recombinación por *crossing-over* se forman una o dos cadenas dobles de ADN heteroduplex entre cromátidas no hermanas; una cadena sencilla proveniente de cada uno de los ADNs parentales. El tamaño de los fragmentos recombinantes es variable porque la estructura heteroduplex puede resolverse de dos maneras diferentes (Holliday, 1964). En la primera, una escisión horizontal, paralela a las cromátidas, genera dos dobles cadenas con un fragmento de recombinación idéntico al tamaño de la hebra heteroduplex pero que conservan el resto del material genético intacto (fenómeno también llamado conversión genética sencilla). En la segunda, un corte vertical hace que se intercambien las cadenas de ADN de las cromátidas a partir del punto original donde se formó la cadena heteroduplex, produciendo grandes fragmentos de recombinación (también conocido como conversión genética con recombinación).

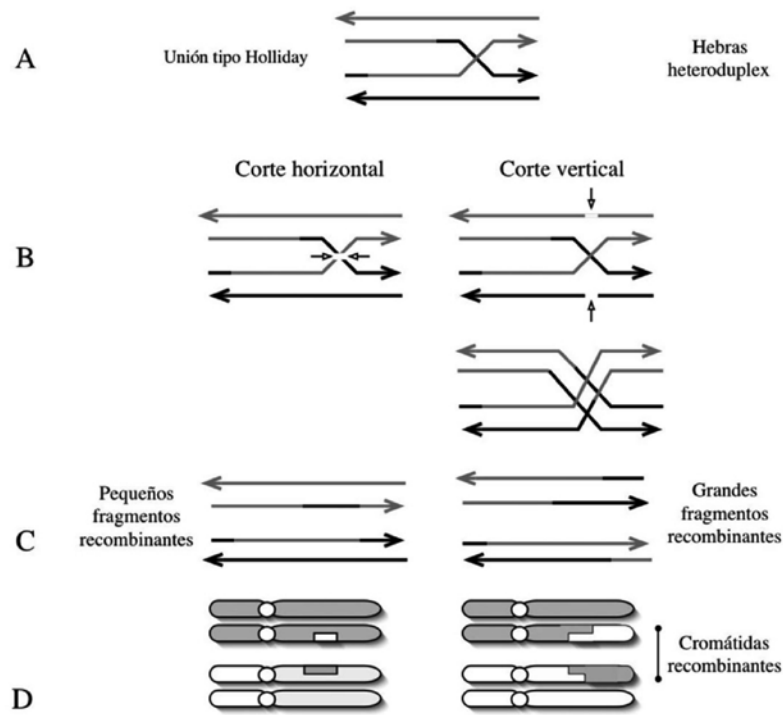
Durante la recombinación por *crossing-over* en eucariontes pueden ocurrir dos fenómenos de importancia evolutiva. Uno es la conversión de genes, que ocurre cuando no se da intercambio recíproco de ADN, sino el reemplazamiento de una de las secuencias parentales por la otra, produciendo dos copias

Figura 1. Recombinación por arreglos independientes durante la meiosis. (A) Los alineamientos diferentes se dan por el giro diferencial de los cromosomas antes de la mitosis I, (B) de tal manera que cada una de las alineaciones (C) producirá gametos con diferentes genotipos. (Modificado de Grifiths *et. al*., 1999)



del mismo locus en lugar de dos copias diferentes (Balding *et al.*, 1992). El otro es el entrecruzamiento desigual, que ocurre cuando hay un corrimiento en la alineación de las secuencias homólogas antes del entrecruzamiento, lo que produce copias extra en una de las cadenas recombinantes y supresiones en la otra. El entrecruzamiento desigual ha jugado un papel muy importante en la evolución de familias multigénicas (*i.e.* por evolución concertada) y en la duplicación de genes, en general (Ohta, 2000).

Figura 2. Recombinación por entrecruzamiento o *crossing over* durante la meiosis. (A) Formación de hebras Heteroduplex entre cadenas de adn de cromátidas no hermanas y uniones tipo holliday. (B) Hay dos maneras de resolver las uniones Holliday, con un corte horizontal o uno vertical que (C) producen fragmentos recombinantes de diferente tamaño, como puede verse en (D) la representación gráfica de las cromátidas recombinantes (modificado de Grifiths *et al*.,1999)



Mecanismos de la recombinación en procariontes

En los procariontes, la recombinación puede ocurrir mediante varios procesos que tienen diferentes consecuencias genéticas. En estos procesos está implicado el reemplazamiento, el intercambio o la inserción de fragmentos de ADN entre diferentes individuos, en fenómenos mejor conocidos como

transferencia lateral de genes (Ochman *et al.*, 2000); aunque también pueden darse rearreglos de ADN dentro del mismo organismo (Alokam *et al.*, 2002; Liu y Sanderson, 1998; Sanderson y Liu, 1998). A diferencia de los eucariontes, los procesos recombinatorios de los procariontes están desligados completamente de la reproducción.

Existen tres mecanismos principales que generan algún tipo de recombinación en procariontes:

1) En la transducción, fragmentos de ADN de un hospedero infectado con el genoma de un virus pueden ser transportados por el mismo virus hasta otro u otros hospederos. Los fragmentos pueden haber sido tomados al azar del genoma del hospedero o sólo haber tomado aquellos fragmentos adyacentes al lugar de inserción del genoma viral. En cualquiera de los dos casos, el resultado es la inserción de secuencias foráneas en un hospedero secundario. La maquinaria del virus protege y ayuda a la inserción del material genético transducido por él. La obtención de genomas completos de varios procariontes ha revelado la importancia de este proceso en la plasticidad genómica de algunos patógenos (Dobrindt y Hacker, 2001).

2) En la transformación, una bacteria puede tomar ADN desnudo del ambiente e incorporarlo a su genoma, sea el central o como plásmido. Existen bacterias que son capaces de este proceso de forma ininterrumpida, como *Neisseria gonhorreae* y *Haemophilus influenzae*, otras que sólo cuentan con la capacidad de incorporar ADN foráneo en ciertas etapas de su ciclo de vida, como *Bacillus subtilis* y *Streptococcus pneumoniae* (Dubnau, 1999) y otras más que pueden ser inducidas de manera artificial, como *Escherichia coli*. Sin embargo, existen limitantes al ADN foráneo que puede insertarse finalmente al genoma: a) las enzimas de restricción del receptor, que distinguen secuencias específicas de ADN, pueden eliminar cualquier fragmento de ADN que no sea compatible con el genoma de éste y b) existen secuencias de transformación (*uptake sequences*) que limitan la formación de hebras heteroduplex, evitando la recombinación (Lorenz y Sikorski, 2000; Majewski y Cohan, 1998; Saunders *et al.*, 1999).

3) Durante la conjugación, una bacteria donadora transfiere fragmentos de ADN de diverso tamaño, del genoma central o plásmidos, a una bacteria receptora, produciendo fragmentos recombinantes de muy diversos tamaños. Al igual que en la transformación, este fenómeno depende generalmente de la distancia filogenética entre donador y receptor, aunque se han observado transferencias de este estilo incluso entre dominios (Heinemann y Sprague, 1989).

Consecuencias de la recombinación

La recombinación *sensu lato* puede llevarse a cabo por mecanismos muy diferentes, dependiendo del sistema de que se trate; sin embargo, las consecuencias evolutivas de la recombinación son las mismas en todos los casos (Posada *et al.*, 2002). Por un lado, genera nuevas combinaciones de alelos parentales en los genotipos (Milgroom, 1996), ya sea por entrecruzamiento de secuencias homólogas o mezclando la diversidad genética entre cromosomas por rearreglos independientes o reemplazamiento. Asimismo, rompe la correlación genealógica entre diferentes regiones de un genoma; es decir, el ligamiento entre los genes (McVean *et al.*, 2002).

A su vez, la generación de nuevos genotipos por recombinación permite la dispersión de múltiples mutaciones potencialmente benéficas (Fisher, 1932; Müller, 1932) y evita la acumulación de mutaciones deletéreas (Müller, 1964); mientras que, al generar independencia entre diferentes regiones del genoma, permite que fuerzas como la selección natural o la deriva genética actúen también de manera independiente sobre los diferentes alelos.

En los eucariontes, la recombinación está ligada a la reproducción sexual, predominante en animales y plantas, por lo que la recombinación está involucrada directamente con la función de diseminar alelos homogéneamente en la población, reduciendo la varianza genética y dándole cohesividad histórica al linaje en donde ocurre (Maynard-Smith *et al.*, 2000). Por el contrario, en los procariontes el intercambio genético no está limitado a la reproducción ni confinado a un linaje determinado, por lo que su efecto es diluir la cohesividad histórica dentro de un linaje (Maynard-Smith *et al.*, 2000), creando filogenias en forma de red (véase el capítulo 4 de este libro).

En el nivel molecular, la recombinación produce "mosaicos" de genes con diferentes historias a lo largo del genoma. Estos patrones son poco evidentes cuando la recombinación ocurre entre fragmentos parecidos y es fácil de identificar cuando se produce entre secuencias más divergentes. Los diferentes detectores y estimadores de la recombinación se sirven, de una u otra forma, de estos patrones para inferirla. Sin embargo, otras fuerzas que modifican los patrones de variación genética, como la mutación, la selección natural y la deriva, pueden interferir en mayor o menor escala sobre esta inferencia, sin que, hasta ahora, se haya estudiado formalmente el problema (Posada *et al.*, 2002).

A su vez, la recombinación afecta potencialmente cualquier inferencia basada en los supuestos de la filogenia tradicional. Al independizar diferentes

Figura 3. La recombinación puede generar topologías ambiguas. Arriba, topologías esperadas bajo los efectos de la recombinación. Abajo, linajes con eventos recombinatorios marcados con flechas horizontales. (A) topología resuelta de un linaje clonal. (B) un evento recombinatorio causa ambigüedades en una de las ramas. (C) dos eventos de recombinación son suficientes para nublar la historia del linaje

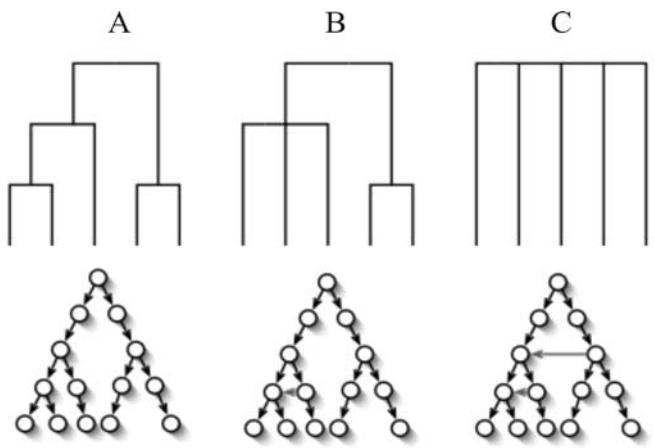
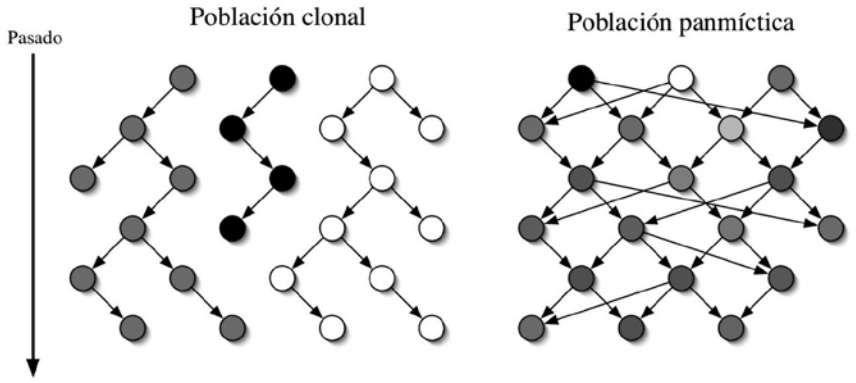


Figura 4. Poblaciones clonales vs. poblaciones panmícticas. Las poblaciones clonales conservan el mismo genotipo a lo largo del tiempo; mientras que las poblaciones panmícticas generan nuevos genotipos por combinación de los alelos de la población



regiones del genoma, la recombinación permite que tengan diferentes historias y, como consecuencia, se producen en los análisis muchas homoplasias o incongruencias filogenéticas (figuras 3 y 4). De tal manera que, si se aplica la inferencia filogenética desconociendo la existencia de la recombinación, la heterogeneidad de tasas de substitución se sobreestimará, la hipótesis del reloj molecular se rechazará falsamente y la longitud de las ramas terminales de un árbol será sobreestimada, produciendo dataciones mal interpretadas (Posada, 2001; Posada y Crandall, 2002; Posada *et al.*, 2002; Schierup y Hein, 2000a; 2000b). Un efecto similar puede esperarse en pruebas que buscan detectar la selección natural (Anisimova *et al.*, 2003) y en estudios que utilizan la coalescencia tradicional para inferir parámetros de demografía histórica u obtener estimados basados en sus simulaciones (Nordborg, 2001).

Las poblaciones microbianas: clonalidad vs. panmixia

Los microorganismos -hongos y bacterias haploides- tienen formas diferentes de reproducirse e intercambiar material genético en comparación con los organismos macroscópicos, lo que hace que la dinámica de sus poblaciones también sea diferente.

Los microbios son capaces de alternar ciclos reproductivos sexuales y asexuales e incluso no presentar ciclos sexuales, dependiendo de una multitud de factores ecológicos y genéticos, que pueden variar en el tiempo y el espacio (Anderson y Kohn, 1998). Debido a esto, sus poblaciones pueden encontrarse en cualquier punto entre la panmixia y la clonalidad (figura 4).

En una población clonal, un mismo genotipo es replicado en muchos individuos y sólo puede surgir uno nuevo con la acumulación de mutaciones. La selección natural actúa sobre los individuos, no sobre sus variantes alélicas, haciendo que las mutaciones exitosas sean fijadas rápidamente y las mutaciones letales purgadas. Sin embargo, fenómenos como el *hitchhiking* (aventón) ocurrirían a nivel genómico (Grigg y Suzuki, 2003; Halliday y Carter, 2003), por lo que se puede mantener una carga genética muy alta y, al mismo tiempo, perderse diversidad genética.

En cambio, en una población panmíctica haploide cada individuo posee un genotipo diferente. La selección natural actúa sobre loci independientes, no sobre los individuos; además, al no haber alelos recesivos en organismos haploides, se puede purgar la carga genética con relativa continuidad. De la misma manera, el constante rearreglo de alelos puede generar genotipos

exitosos más efectivamente que en las poblaciones clonales; aunque puede romperlos con la misma facilidad. Asimismo, los fenómenos como el *hitchhicking* sólo pueden ocurrir en genes ligados.

Las poblaciones con estructuras intermedias entre la panmixia y la clonalidad generalmente presentan ciclos epidémicos, en donde existe la recombinación pero es opacada por la reproducción asexual a gran escala, sobre representando en la muestra sólo algunas clonas. Estas poblaciones pueden tener cantidades muy variables de genotipos diferentes, dependiendo de la frecuencia con que ocurra la recombinación y la dispersión de individuos de origen puramente clonal. Sin embargo, un solo evento de recombinación libre es suficientemente poderoso para romper ligaduras entre genes, purgar alelos deletéreos y generar nuevos genotipos en una población (Milgroom, 1996), por lo que aún en estas poblaciones la recombinación regirá la evolución en el largo plazo (Maynard-Smith *et al.*, 1993).

Cabe señalar que la diversidad genotípica es independiente de la diversidad genética; es decir, el hecho de tener muchos genotipos diferentes no garantiza un número alto de alelos diferentes para cada locus y tener un número bajo de genotipos en una población no implica necesariamente que la variación genética también sea baja (ver capítulo 3). Por ejemplo, *Pseudomonas stutzeri*, causante de gastroenteritis, posee una variación genética muy alta ($H = 0.876$) que no es explicada por la presencia de la recombinación solamente (Rius *et al.*, 2001). En general, la recombinación no genera alelos nuevos, sólo los acomoda en diferentes combinaciones en un genoma, es independiente de la variación genética vista como número de alelos por loci.

Ejemplos de recombinación en bacterias

La ocurrencia de la recombinación en bacterias se demostró experimentalmente hace más de cincuenta años (Ledeerberg y Tatum, 1946). Sin embargo, aún está por descubrirse el papel que juega en las poblaciones naturales y en los complejos de bacterias comensales y patógenas que han sido los más estudiados (Maynard-Smith *et al.*, 2000).

Existen bacterias que son estrictamente clonales como *Borrelia*, causante de la artritis de Lyme (Dykhuizen *et al.*, 1992), y bacterias de estructura altamente recombinante, como *Helicobacter pylori*, responsable de úlceras estomacales (Suerbaum *et al.*, 1998). La mayoría de las bacterias se encuentra en cualquier punto entre estos dos extremos. De hecho, aún en las bacterias en donde la

recombinación domina la evolución a largo plazo, éstas no pueden escapar de la aparición repentina de clonas de amplia dispersión o epidémicas (Maynard-Smith *et al.*, 2000).

No siempre es sencillo distinguir una población con estructuras intermedias entre la clonalidad y la panmmixia porque existen muchas variantes. Los casos más estudiados se asocian a la dimensión temporal, es decir, que la recombinación produzca en un tiempo dado una clona exitosa que se dispersa eficazmente y luego desaparezca. Por ejemplo, *Porphyromonas gingivalis*, causante de infecciones periodontales, posee clonas ventajosas de distribución mundial recurrente, aunque se ha detectado que la recombinación es frecuente (Frandsen *et al.*, 2001). Un poco menos estudiados son los casos en los que las estructuras poblacionales intermedias se deben a la dimensión espacial, en donde la recombinación no parece verse coartada por el surgimiento de clonas exitosas, sino por la geografía. Por ejemplo, el género *Nodularia* de cianobacterias tiene recombinación limitada geográficamente; cuando se analizan todos los representantes de *Nodularia* en el Báltico, parecería que se trata de poblaciones clonales, pero cuando se subdivide la población con base en la geografía es posible detectar recombinación entre ellas (Barker *et al.*, 2000). Un patrón similar de aislamiento pudo observarse en *Rhizobium leguminosarum* bv. *phaseoli,* que presenta una aparente estructura clonal cuando se analiza Norteamérica como un solo grupo de datos, pero cuando se analizan parcelas por separado se encuentra evidencia de recombinación (Souza *et al.*, 1992). Otro tipo de aislamiento que da estructuras intermedias se presenta en otro rizobio, *Rhizobium etli* bv. *phaseoli*, donde se evidenció una estructura recombinante dentro de cada uno de dos linajes, pero no entre ellos (Silva *et al.*, 1999), lo que hace pensar sobre una definición biológica de especie basada en los límites al intercambio genético (véanse los capítulos 10 y 11 de este libro).

También se han estudiado los alcances de la recombinación en el genoma de algunas bacterias. Por ejemplo, se sabe que en el meningococo *Neisseria meningitidis* el 50% de los alelos surgidos por recombinación están presentes en cualquier otra parte de la población, mientras que en el pneumococo *Streptococcus pneumoniae* es el 80% (Feil *et al.*, 2000a). Los fragmentos de ADN recombinante son de 5 a 10 kb en ambos microorganismos, pero en el primero, el intercambio de alelos se da en secuencias que divergen en promedio un 4%, mientras que en el segundo es un 1%, lo que sugiere que existe intercambio genético entre diferentes especies de *Neisseria* (Feil *et al.*, 2001). Además, en *Neisseria* pueden detectarse estructuras de mosaico que revelan

la recombinación entre genomas diferentes, probablemente de especies diferentes, y también zonas de alta variación que sugieren que la recombinación se da también entre individuos parecidos, probablemente de la misma especie (Feil *et al.*, 2001; Feil *et al.*, 2000a).

Es posible que lo observado en *Neisseria* sea un patrón generalizable a otras bacterias patógenas con recombinación recurrente. Sin embargo, la recombinación, seguida de la propagación de clonas exitosas no siempre es la mecánica de los patógenos. Se sabe que las cepas de *Staphylococcus aureus* que causan infecciones sistemáticas en la tracto respiratorio no son clonas de reciente aparición, provenientes de las cepas de pacientes sanos (Feil *et al.*, 2003), sino que son las mismas, cuya mutación mantiene una variación suficientemente alta como para enfrentar las defensas del hospedero y los tratamientos con antibióticos. De hecho, la contribución de la mutación a la variación en este organismo es 15 veces la aportada por la recombinación, lo que contrasta con *Neisseria meningitidis*, *Streptococcus pneumoniae* y *Campylobacter jejuni*, en donde la recombinación es, respectivamente, 5, 10 (Feil *et al.*, 2003) y 50 (Schouls *et al.*, 2003) veces más importante que la mutación.

Aún en bacterias cuyos loci no recombinan entre sí, parece que la recombinación por transferencia horizontal es un evento recurrente, aunque no sepamos bien a bien sus orígenes. Se ha podido estimar, por ejemplo, que *E. coli* importa ADN foráneo a una tasa de 16 kb por cada millón de años (Lawrence y Ochman, 1998) y que entre el 10 y 16% de su genoma se originó por transferencia horizontal (Médigue *et al.*, 1991), lo que equivale a las diferencias encontradas entre su genoma y el de la patógena *Salmonella enterica* (Ochman *et al.*, 2000).

Ejemplos de recombinación en hongos filamentosos

La recombinación en hongos se puede dar, como en el resto de los eucariontes, por rearreglos independientes o por entrecruzamiento en la meiosis. En pocos casos se pueden producir genotipos nuevos por la segregación de cromosomas enteros, en lo que se conoce como ciclo parasexual (Fincham *et al.*, 1979; Prado *et al.*, 2003), por lo que la producción de genotipos nuevos está casi confinada a la presencia de reproducción sexual.

Los hongos producen una multitud de propágulos mitóticos y meióticos, por lo que la estructura de sus poblaciones es parecida al modelo epidémico de las bacterias. Sin embargo, existen poblaciones aparentemente panmícticas,

como el hongo liquenizado *Letharia* (Kroken y Taylor, 2001) y el patógeno humano *Coccidioides immitis* (Koufopanou *et al.*, 2001). También se han encontrado poblaciones exclusivamente clonales, como *Trychophyton rubrum,* el principal causante de dermatomicosis (Gräser *et al.*, 1999).

Un fenómeno que puede alterar las inferencias sobre recombinación en los hongos es la posibilidad que tienen de autofecundarse, como es el caso de *Cryphonectria parasitica,* en el que parte de la estructura clonal se debe a la capacidad que tiene este organismo de alternar autofecundación con fecundación cruzada y no a la mera producción de esporas asexuales (Milgroom, 1996).

Aún más, existen casos de poblaciones con estructura aparentemente panmíctica en los que no se han reportado estructuras sexuales, como es el caso de *Aspergillus flavus,* un hongo productor de aflatoxinas (Geiser *et al.*, 1998) y *Coccidioides immitis* (Burt *et al.*, 1996; Koufopanou *et al.*, 2001) y no siempre se ha podido identificar el origen de la variación encontrada (Anderson y Kohn, 1998). En el caso particular de *C. immitis*, aún falta profundizar sobre la supuesta estructura panmíctica. Recientemente se ha encontrado que las cepas responsables de la enfermedad no pertenecen a ninguna clona en particular sino a varios complejos diferentes y sin relación aparente. En otras palabras, las epidemias no fueron causadas por clonas hipervirulentas, sino a la disponibilidad de inóculo por el exceso de lluvias e inundaciones anteriores al surgimiento de la epidemia (Fisher *et al.*, 2000); la rápida expansión de la población pudo contribuir con mutaciones a la variación observada, por lo que hace falta distinguir con detalle cuál ha sido la contribución neta de la recombinación.

Por otra parte, aunque algunos hongos han sido utilizados como organismos modelo para estudiar los mecanismos finos de la recombinación (p. ej. Holliday, 1964), se sabe muy poco de la importancia que tiene ésta en poblaciones naturales; por ejemplo, los estudios sobre recombinación se han limitado a su detección, no a la estimación de su frecuencia; esta última, sin lugar a dudas, indicaría nuevas direcciones en el estudio de especies de importancia humana, como *C. immitis* o los *Aspergillus*.

Detectores y estimadores de la recombinación

Dos aspectos que se busca conocer cuando se estudia la recombinación son si ha ocurrido y con qué frecuencia. La detección de la recombinación es de interés para entender la evolución de los genomas bacterianos, para es-

tudiar los modos de reproducción en los hongos (Kauserud y Schumacher, 2001; Koufopanou *et al.*, 2001; Kroken y Taylor, 2001), para comprender la evolución de familias multigénicas (Schierup *et al.*, 2001) y, recientemente, los efectos de la recombinación en la inferencia filogenética (Posada, 2001; Posada y Crandall, 2001a; 2002; Schierup y Hein, 2000a; 2000b). Para detectar recombinación no se necesitan datos poblacionales, se puede estudiar en linajes diferentes y en familias de genes, sean ortólogos o parálogos (véase el capítulo 4 de este libro).

Estimar la frecuencia de la recombinación se refiere a estimar el parámetro de la tasa de recombinación poblacional (ver más adelante). Conocer la frecuencia de la recombinación en una muestra puede ser útil para predecir la dinámica de la población de interés; por ejemplo, qué tan pronto un patógeno se hará resistente a un fármaco, cambiará de hospedero o disminuirá su patogenicidad. En el caso de los hongos, se podría discernir con mayor precisión si la estructura panmíctica observada en especies asexuales se debe a eventos del pasado o a recombinación consuetudinaria independiente del sexo. Para estimar la frecuencia de la recombinación se requieren datos poblacionales, casi siempre secuencias de ADN de genes ortólogos.

Detección de la recombinación

Existen dos aproximaciones diferentes para la detección de la recombinación. La primera, relativamente empírica, se basa en cualidades contingentes de los o partes de los genomas; sirve principalmente para detectar eventos de transferencia horizontal que involucran la inserción de secuencias en un genoma. La segunda, más formal, busca patrones que pudieran ser resultado de la recombinación, comparando generalmente alineaciones de secuencias homólogas; es útil para detectar recombinación tanto entre genes ortólogos como parálogos, en cualquier nivel taxonómico.

Métodos empíricos

Los métodos empíricos para detectar recombinación tratan de reconocer secuencias anormales en relación con otras partes del genoma por su composición nucleotídica (Lawrence y Ochman, 1997); las frecuencias de dinucleótidos (Karlin, 1998); el uso preferencial de codones (Médigue *et al.*, 1991; Moszer *et al.*, 1999; Mrazek y Karlin, 1999) u otros patrones inferidos por cadenas de Markov (Hayes y Borodovsky, 1998). Una vez detectadas estas anomalías

y después de desechar aquellas que por su composición de aminoácidos dan estos patrones, se proponen regiones que hayan sido transferidas lateralmente (Lawrence y Ochman, 1998). En general, estos métodos sólo pueden aplicarse cuando se tienen secuenciadas regiones vastas del genoma o genomas completos y sólo son recomendables cuando se espera que la recombinación sea un fenómeno excepcional y reciente entre linajes diferentes. Además, los patrones de uso preferencial de codones y composición nucleotídica no siempre están relacionados con eventos de recombinación, sino que pueden ser generados por otras razones, como el nivel de expresión de la región analizada (Médigue *et al.,* 1991) y otras aún no bien definidas (Guindon y Perriere, 2001; Koski *et al.*, 2001). Aún los métodos que utilizan como evidencia a ortólogos posicionales, en lugar de establecer homología por similitud (Koski *et al.*, 2001), son poco efectivos cuando han ocurrido rearreglos, inversiones o supresiones dentro de un genoma.

La comparación de los diferentes métodos para detectar transferencia horizontal mostró que sus resultados pueden sobrelaparse pero, en general, son diferentes (Ragan, 2001). Esto se debe a que cada método busca sus propias anomalías en la secuencia basándose en una descripción diferente de lo que es "normal" para un genoma dado (Lawrence y Ochman, 2002). En conclusión, utilizar más de un método es recomendable, aunque la comparación de los resultados debe ser escrupulosa (Ragan, 2002).

Alternativamente, es posible usar como evidencia de transferencia horizontal la presencia de remanentes de elementos translocables, secuencias específicas de plásmidos conjugativos o sitios de unión a integrasas de fagos en las secuencias adyacentes a ellas. Claramente, esta aproximación sólo es útil cuando la recombinación fue originada por conjugación de plásmidos, elementos móviles o transducción.

Métodos formales

Los métodos de detección de la recombinación buscan, de una u otra forma, los patrones de mosaico que ésta produce en el nivel molecular. Por tratarse de métodos que no estiman parámetros relacionados con la tasa de recombinación, se les ha denominado no paramétricos, aunque puede haber diferentes clasificaciones. Posada y Crandall (2001b) y Posada (2002) dividen los métodos no paramétricos en cinco categorías no exclusivas: los métodos de similitud, de distancias, los filogenéticos, de compatibilidad y los de sustitución (tabla 1).

Los métodos de similitud infieren la recombinación cuando las sustituciones sinónimas en regiones variables exceden aquellas en regiones conservadas. Estos métodos no han sido usados extensivamente, en parte porque sólo son útiles en la detección de conversión génica en familias de genes y sólo pueden aplicarse a regiones codificantes (Feil *et al.*, 2001; Marais *et al.*, 2001; Menotti-Raymond *et al.*, 1991; Ohta y Basten, 1992). La selección natural o el uso preferencial de codones podrían dar patrones que confundan esta inferencia.

Los métodos de distancia buscan inversiones de patrones de distancia entre secuencias (Weiller, 1998). En general, usan una aproximación de ventanas deslizantes (*sliding windows* en inglés) y la estimación de algunos estadísticos basados en distancias genéticas entre las secuencias. Como la filogenia no necesita saberse, estos análisis son rápidos.

Los métodos filogenéticos infieren la recombinación cuando las filogenias de diferentes partes del genoma poseen topologías discordantes o cuando genes ortólogos de especies diferentes se agrupan. Se puede inferir la recombinación si la comparación de secuencias adyacentes proporciona diferentes patrones de ramificación y, más aún, cuando la consecuencia de tales cambios resulta en la conciliación de diferentes filogenias de secuencias en una sola filogenia. Estos son los métodos más usados en la literatura (Balding *et al.*, 1992; Fitch y Goodman, 1991; Grassly y Holmes, 1997; Hein, 1990; 1993; Holmes *et al.*, 1999; Husmeier y Wright, 2001; Koop *et al.*, 1989; Martin y Rybicki, 2000; McGuire *et al.*, 1997; McGuire y Wright, 1998; 2000; McGuire *et al.*, 2000; Robertson *et al.*, 1995a; Robertson *et al.*, 1995b; Salminen *et al.*, 1996; Siepel y Korber, 1995; Siepel *et al.*, 1995).

Para entender mejor los límites de los métodos filogenéticos, Wiuf *et al.* (2001) distinguen tres diferentes escenarios de la recombinación sobre una filogenia. En el primero, dos linajes recombinantes coalescen entre ellos antes de hacerlo con cualquiera no recombinante; en este caso, no hay ninguna alteración en la topología de la filogenia y tampoco la hay en la longitud de las ramas: el evento de recombinación es indetectable (figura 5 A y B). En el segundo escenario, uno de los linajes recombinantes coalesce con uno no recombinante antes de hacerlo con el otro; en este caso, la topología no se altera, pero sí las longitudes de las ramas, de tal manera que, si el método en turno puede reconocer este tipo de incongruencias, el evento recombinatorio será detectable (figura 5 D y E). La tercera posibilidad es que al menos uno de los linajes recombinantes coalesca con dos o más linajes no recombinantes antes de hacerlo con el otro, resultando en una topología distinta para cada una de

Tabla 1. Programas detectores de la recombinación (modificada de Posada, 2002)

Programa	Tipo	P[a]	Par[b].	Qui[c].	Mideρ	P_E[d]	P_O	E_E[e]	E_O	Referencia
Chimaera	Sustituciones	Permutaciones	Sí	Sí	No	0.69	0.75	0.03	0.08	Posada y Crandall, 2001
Geneconv	Sustituciones	Permutaciones	Sí	Sí	No	0.67	0.92	0.04	0.25	Sawyer, 1999
Prueba de homoplasia	Sustituciones	Permutaciones	No	No	Sí	0.23	0.36	0.06	0.2	Maynard-Smith y Smith, 1998
MaxChi2	Sustituciones	Permutaciones	Sí	Sí	No	0.69	0.92	0.03	0	Maynard-Smith, 1992
PhyPro	Distancias	Permutaciones	Sí	Sí	No	0.68	0.58	0.02	0.08	Weiller, 1998
Pist	Sustituciones	Monte Carlo	No	No	Sí	0.58	0.82	0.01	0.42	Worobey, 2001
Plato	Filogenia	Monte Carlo	No	Sí	No	0.39	0.7	0.04	0.25	Grassly y Holmes, 1997
Rdp	Filogenia	Distribución binomial	Sí	Sí	No	0.63	0.83	0.01	0	Martin y Rybicki, 2000
RecPars	Filogenia	Sí/No	Sí	Sí	No	0.68	0.45	0.04	0.08	Hein, 1990
Reticulate	Compatibilidad	Permutaciones	No	No	No	0.67	0.91	0.03	0.27	Jakobsen y Easteal, 1996
Prueba de corridas	Sustituciones	Distancia geométrica	No	No	Sí	0.19	0.58	0.06	0.5	Takahata, 1994
Simplot	Filogenia	Sí/No	Sí	Sí	No	0.43	0.27	0.01	0.8	Salminen *et al.*, 1996; Lole *et al.*, 1999

(Continúa)

Tabla 1. Programas detectores de la recombinación (modificada de Posada, 2002)

Programa	Tipo	P[a]	Par[b].	Qui[c].	Mideρ	P_E[d]	P_O	E_E[e]	E_O	Referencia
Prueba de Sneath	Sustituciones	Dist. Normal	Sí	No	No	0.67	0.92	0.04	0.33	Sneath, 1995
Triple	Filogenia	Dist. Derivada	Sí	Sí	No	0.68	0.75	0.02	0.42	Kuhner *et al.*, 1991

a. Se refiere a cómo se obtiene un valor de confianza P. b. El método evidencia posibles secuencias parentales. c. El método encuentra puntos de quiebre de la recombinación. d. Poder esperado (PE) y poder observado (P_O): el poder esperado es el promedio (ponderado con la composición de las secuencias empíricas de la tabla 2) del porcentaje de las veces que un método detectó recombinación cuando la mayoría de los otros lo hicieron ($R > 0.5$) para diferentes diversidades ($\theta = 10, 50, 100$ y 200) y niveles de recombinación ($\rho = 1, 4, 16, 64$), sin heterogeneidad de tasas de substitución entre los sitios ($\alpha = \infty$), donde las secuencias simuladas fueron obtenidas con el modelo de dos parámetros de Kimura (1980) por Posada y Crandall (2001b); el poder observado es el porcentaje de las veces que un método en particular detectó recombinación cuando la mayoría de los otros lo hicieron ($R > 0.5$), para los datos empíricos de la tabla 2. e. Error (falsos positivos) esperados (EE) y observados (EO): el error esperado es el promedio (ponderado con la composición de las secuencias empíricas de la tabla 2) de las veces que un método detectó incorrectamente recombinación, con diversos valores de heterogeneidad de tasas de substitución entre sitios ($\alpha = 0.05, 0.5, 2, \infty$) y diversidad ($\theta = 10, 50, 100, 200$), sin recombinación ($\rho = 0$) donde las secuencias simuladas fueron obtenidas con el modelo de dos parámetros de Kimura (1980) por Posada y Crandall (2001b); el error observado es el promedio de las veces que un método detectó recombinación cuando la mayoría no lo hizo ($R < 0.5$), con los datos empíricos de la tabla 2.

las regiones a los lados del punto de quiebre (figura 5-C). Este tipo de eventos son los que mejor pueden detectarse por los métodos filogenéticos.

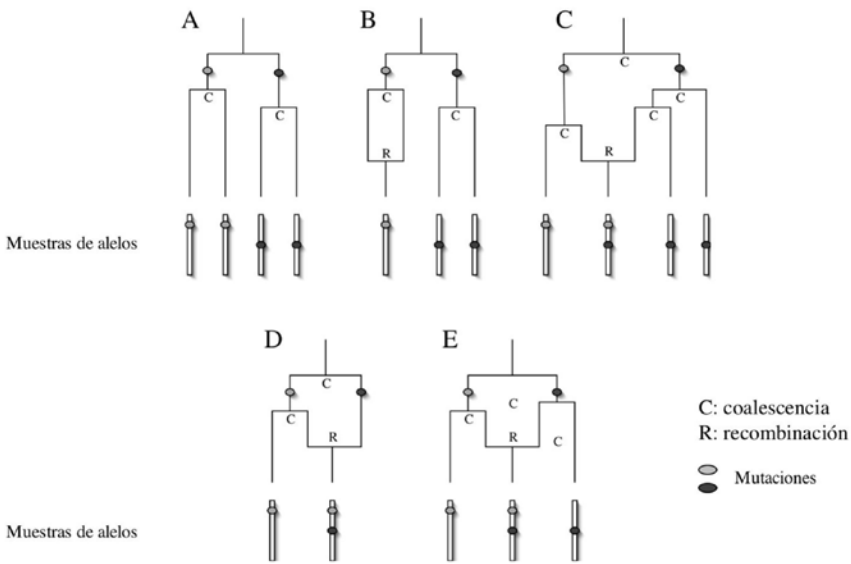
Por su parte, los métodos de compatibilidad tratan de probar la existencia de particiones en la incongruencia filogenética basados en un análisis sitio por sitio; no requieren conocer la filogenia de las secuencias analizadas (Drouin y Dover, 1990; Jakobsen y Easteal, 1996; Jakobsen *et al.*, 1997; Sneath *et al.*, 1975).

Los métodos de distribución de sustituciones examinan las secuencias para encontrar agrupaciones significativas de sustituciones o para que cumplan una distribución estadística esperada (Betrán *et al.*, 1997; DuBose *et al.*, 1988; Gibbs *et al.*, 2000; Imanishi, 1996; Maynard-Smith, 1992; Maynard-Smith y Smith, 1998; McVean *et al.*, 2002; Satta, 1992; Sawyer, 1989;1999; Sneath, 1995;1998; Stephens, 1985; Takahata, 1994; Valdés y Piñero, 1992; Worobey, 2001).

El desequilibrio de ligamiento

Un caso particular de detección de la recombinación es el desequilibrio de ligamiento (DL; Lewontin y Kojima, 1960). Se diferencia del resto de los

Figura 5. Tres diferentes escenarios de la recombinación sobre una filogenia (Wiuf *et al.* 2001)



detectores de la recombinación porque necesariamente utiliza datos multilocus (véase el capítulo 18 de este libro) y porque está relacionado con otros parámetros de genética de poblaciones de una manera directa, por lo que, en un contexto apropiado, también puede utilizarse de una manera paramétrica (ver Feil *et al.*, 2000b).

Calcular el desequilibrio es ver qué tan asociados están los alelos de varios loci dentro de una misma muestra (Gaut y Long, 2003). Si la población es completamente clonal, siempre se presentarán los mismos alelos de varios loci juntos; mientras que si es completamente panmíctica, se observarán todas las posibles combinaciones de alelos.

Existen muchos métodos para medir el DL, todos basados en los principios antes mencionados y divergentes sólo por la solución a problemas particulares. El DL merecería *per se* una revisión completa (p. ej. Gaut y Long, 2003), por lo que sólo se revisará el indicador más usado entre los microbiólogos.

La medida más popular del DL es el índice de asociación I_A de Maynard-Smith *et al.* (1993), que es una versión mejorada de los índices de Brown *et al.* (1980) y Whittman (1983). En ella se utilizan las frecuencias alélicas de varios loci para establecer si todas las combinaciones de alelos están presentes en una muestra. El índice de asociación se obtiene al calcular la diferencia entre la varianza que se esperaría si todos los alelos estuvieran distribuidos aleatoriamente y la varianza observada de esta distribución. Si no hay diferencia significativa entre las varianzas ($I_A = 0$), se puede suponer que la recombinación es libre en la muestra. Los valores diferentes de cero deben ser tratados con cuidado pues pueden significar que la recombinación está ausente, pero también que está restringida a linajes aislados por la geografía, el hospedero, etc. (Maynard-Smith *et al.*, 1993; Souza *et al.*, 1992).

Se han hecho variaciones a este índice para calcular la varianza de las diferencias pareadas sin asumir una distribución normal, obteniendo una distribución exacta por simulaciones de Monte Carlo (Haubold *et al.*, 1998; Haubold y Hudson, 2000), y una diferente estandarización de las covarianzas, que elimina la dependencia del índice del tamaño de la muestra (Agapow y Burt, 2001).

Cabe señalar que el DL es influenciado por factores históricos, diferentes de la recombinación, que afectan el tamaño efectivo de la población (ver capítulo 3). Así, la subdivisión de poblaciones, los cuellos de botella y la selección direccional suelen aumentar el DL; aunque sus efectos pueden ser transitorios (Przeworski, 2002; Wall *et al.*, 2002). El sistema de apareamiento también tiene efectos: las especies que se autofecundan tienen tasas de recombinación por

meiosis más altas que las que se entrecruzan (véase el capítulo 6 de este libro; Charlesworth y Charlesworth, 1979).

El tipo de datos que utiliza el DL no permite su comparación con los otros métodos de detección, por lo que no será incluido en la siguiente sección.

Por otra parte, David Robertson de la Universidad de Manchester, creó una página con enlaces a la mayor parte de los programas que detectan o estiman la recombinación y que se mencionan en este capítulo: www.bioinf.man.ac.uk/~robertson/recombination.

Comparación de los métodos no paramétricos

Recientemente se han publicado comparaciones de varios métodos no paramétricos para detectar recombinación (Brown *et al.*, 2001; Maynard-Smith, 1999; Posada y Crandall, ; Posada, 2002; Wiuf *et al.*, 2001). Posada y Crandall (2001b) y Posada (2002) hacen la comparación más completa al utilizar catorce de los métodos con datos de simulaciones (Posada y Crandall, 2001b) y datos empíricos (Posada, 2002), obteniendo resultados que concuerdan con las otras comparaciones: los métodos varían ampliamente en su grado de certeza y su capacidad de errar (dar falsos positivos), dependiendo de varios factores.

En primer lugar, son mutuamente excluyentes los métodos capaces de detectar eventos raros de recombinación y los que son sensibles a una recombinación recurrente (Maynard-Smith, 1999). Muchos de ellos funcionan bien cuando se espera que la recombinación sea rara o cuando se quiere probar la transferencia horizontal de genes entre linajes separados filogenéticamente y pueden no ser útiles cuando la recombinación es muy común o ha ocurrido hace mucho tiempo, si existen altas tasas de mutación o si algunos de los loci estudiados están sujetos a selección natural o deriva genética (Maynard-Smith, 1999).

La comparación de los métodos con datos reales es bastante ilustrativa de sus capacidades y limitaciones. Posada (2002) compara los resultados de 14 métodos (tabla 1) para detectar recombinación en 24 grupos de datos, 12 de los cuales son presumiblemente recombinantes y los 12 restantes no; ambos provenientes de diversos organismos y de niveles taxonómicos variados (tabla 2). Como no se sabe si realmente hay recombinación en los datos reales, Posada (2002) genera un valor arbitrario R, que es un consenso de detección de los 14 métodos sobre un grupo de datos dado. Asimismo, utiliza R para definir el desempeño de cada método, es decir, cuándo un método fue capaz de detectar la recombinación cuando la mayoría ($R > 0.5$) de los otros lo hicieron (poder

Tabla 2. Variación genética y resumen de los resultados de recombinación para cada locus (modificada de Posada, 2002).

Datos[a]	Gen	Homología	Nivel Tax.	Diver.[b]	θ^c	α^d	R^e	R_s^f
Armillaria-mtDNA	mtDNA	Ortólogo	Especie	0.0022	10	0.002	0.23	0.32
BoletalesATP6	ATP6	Ortólogos	Órden	0.1704	100	0.361	**0.76**	**0.8**
Candida-mtDNA	mtDNA	Ortólogo	Especie	0.007	10	0.152	**0.86**	0.32
Candidula16S	16S	Ortólogos	Especie	0.0187	10	1.469	0.48	0.32
DaphniaCOI	COI	Ortólogos	Género	0.163	50	0.010	0.15	**0.8**
DmelCytB	CytB	Ortólogos	Especie	0.0009	10	∞	0.00	0.32
Fusarium3	Al-Tri101-Pp	Ortólogo	Especie	0.0083	10	∞	**0.79**	0.32
FusariumTri101	Tri101	Ortólogo	Especie	0.0105	50	∞	0.04	**0.8**
GymnND4	ND4	Ortólogos	Especie	0.1837	200	1.308	0.10	**0.83**
HGVgenome	Genoma	Ortólogo	Especie	0.0936	200	0.509	**0.77**	**0.83**
HIV(B)EnvNR	Env	Ortólogos	Subtipo	0.0888	200	0.679	**0.61**	**0.83**

(Continúa)

Tabla 2. Variación genética y resumen de los resultados de recombinación para cada locus (modificada de Posada, 2002) *(continúa)*

Datos[a]	Gen	Homología	Nivel Tax.	Diver.[b]	θ[c]	α[d]	R[e]	R_s[f]
HIVEnv	Env	Ortólogo	Grupo	0.2121	200	0.566	**0.94**	**0.83**
HIVEnvR	Env	Ortólogos	Grupo	0.25456	200	0.519	**0.71**	**0.83**
HumanDRB1	DRB1	Ortólogo	Especie	0.1046	50	∞	**0.77**	**0.8**
HumanHRV1	HRV1	Ortólogos	Especie	0.011	10	∞	0.09	0.32
InsectaCOII	COII	Ortólogos	Clase	0.248	200	6.683	0.23	**0.83**
MaizACT	Actina	Parálogo	Especie	0.1708	200	0.280	**0.71**	**0.83**
MammPDH	Pdh	Parálogo	Clase	0.1889	200	0.441	0.21	**0.83**
MammPGK	Pgk	Parálogo	Clase	0.2294	200	0.473	0.18	**0.83**
NeisseriaArgF	ArgF	Ortólogo	Género	0.1152	50	0.686	**0.71**	**0.8**
Perom12S	12S	Ortólogos	Subfamilia	0.0452	50	0.091	0.25	**0.8**
PetuniaS-Rnase	S-RNAsa	Ortólogo	Especie	0.3604	200	2.385	**0.51**	**0.83**

(Continúa)

Tabla 2. Variación genética y resumen de los resultados de recombinación para cada locus (modificada de Posada, 2002) *(continúa)*

Datos[a]	Gen	Homología	Nivel Tax.	Diver.[b]	θ[c]	α[d]	R[e]	R_s[f]
VertebCOI	COI	Ortólogos	Superclase	0.2292	200	1.281	0.32	**0.83**
WolfCR	Reg. Cont.	Ortólogo	Especie	0.0243	10	0.001	0.18	0.32

a. Las referencias de los datos pueden encontrarse en Posada (2002); en negritas, los datos putativamente recombinantes. b. La divergencia se refiere al promedio de diferencias pareadas por sitio. c. Diversidad θ aproximada para cada grupo de datos, tomando en cuenta el número de sitios y divergencia. d. α es un estimado de heterogeneidad de tasas entre sitios, los números más pequeños corresponden a alta heterogeneidad, mientras que $\alpha = \infty$ significa que no la hay. e. R = número de métodos que infirieron recombinación con una $P < 0.01$*6 + los métodos que lo hicieron con una $P > 0.01$, pero menor a 0.05*3 + los que lo hicieron con una $P > 0.05$, pero menor de 0.1; todo entre 24*6. En negritas, lo que para este índice sería recombinante ($R > 0.5$). f. Rs es el promedio de las veces que los programas detectaron recombinación para una diversidad dada (θ = 10, 50, 100, 200) en los datos simulados por Posada y Crandall (2001b) con un nivel de recombinación de $\rho = 16$, sin heterogeneidad de tasas y generados bajo el modelo de dos parámetros de Kimura (1980); en negritas, lo que para este índice sería recombinante ($Rs > 0.5$).

observado) y cuándo dio falsos positivos (error observado), también a juicio de la mayoría de los métodos ($R < 0.5$). En la tabla 2 se muestran estos valores para cada locus, además de un valor de R_s, que es un consenso arbitrario análogo al de Posada (2002), obtenido a partir de los resultados de las comparaciones con simulaciones de Posada y Crandall (material suplementario, 2001b), para un grupo de datos aproximado a la diversidad de las datos reales, para una tasa de recombinación ρ ($\rho = 4N_e r$, ver apartado de estimación de la recombinación) de 16 y donde las secuencias simuladas fueron generadas con el modelo de dos parámetros de Kimura (1980), sin heterogeneidad de tasas entre los sitios. Esta comparación muestra que se puede inferir erróneamente que no hay recombinación haciendo uso de todos los métodos disponibles a la vez y que los datos reales pueden tener combinaciones de divergencia, heterogeneidad de tasas y recombinación que superan por mucho las expectativas con datos simulados.

De acuerdo con estos resultados, la comparación de los métodos con datos reales mostró que, en general, los métodos son sensibles a la divergencia entre las secuencias; tienen menos certeza cuando se trata de divergencias muy bajas (< 5%, con datos simulados) y dando algunos falsos positivos con divergencias muy altas ($\theta \geq 100$). Asimismo, la heterogeneidad de tasas dentro de los sitios puede generar falsos positivos en algunos programas (*i.e.* la prueba de Homoplasia, RecPars, Rdp y Triple), exacerbándose con la divergencia. Sin embargo, no se encontró una relación entre el número de secuencias, sitios, sitios informativos o el uso de diferentes modelos de sustitución y el poder de los diferentes programas; tanto con datos simulados como con datos reales (Posada, 2002). De la misma manera, la mayor parte de los métodos infiere falsos positivos de una manera aceptable; por debajo de 5% en simulaciones y 26% en promedio (*NB* al promediar la respuesta de todos los programas con el parámetro *R*, se está subestimando la presencia de la recombinación y sobrestimando el error, con datos reales).

Los resultados de Posada y Crandall (2001b) junto con las inferencias de Brown *et al.* (2001) y Wiuf *et al.* (2001) confirman que los métodos de sustitución y de compatibilidad son los mejores para la mayor parte de los casos. Sin embargo, es una buena idea estimar la divergencia entre las secuencias y la heterogeneidad de tasas, ya que si la divergencia y la heterogeneidad son bajas, la prueba de homoplasia (Maynard-Smith y Smith, 1998) es una buena opción; mientras que, si la divergencia es del 5 al 20%, esta misma prueba pierde todo su poder; mientras que Maximum Chi Square (Maynard-Smith,

1992), la prueba de perfiles filogenéticos (PhyPro, Weiller, 1998), Reticulate (Jakobsen y Easteal, 1996) y GENECONV (Sawyer, 1989;1999), son las más apropiadas y no causan demasiados falsos positivos.

Estimación de la recombinación

Métodos paramétricos

Los métodos paramétricos estiman la tasa de recombinación poblacional (C o ρ), en donde $\rho = 4N_e r$, N_e el tamaño efectivo de la población y r la tasa de recombinación por locus por generación, relacionada con θ, en donde $\theta = 4N_e\mu$ (μ es la tasa de mutación por locus por generación); es decir, se intenta discernir entre la variación generada por la mutación (μ) y aquella de origen recombinante (r). Como ambos estimadores dependen del tamaño efectivo de la población, la relación entre θ y ρ puede escribirse como $\varepsilon = \rho / \theta = r / \mu$; siendo ε la razón entre la tasa de recombinación y la tasa de mutación, ambas por locus por generación.

Se han empleado dos aproximaciones para la estimación de ρ. Empíricamente, se observa la frecuencia del intercambio de secuencias entre marcadores distantes, haciendo mapeo genético (p. ej. Ashburner, 1989; Bouffard *et al.*, 1997; Nagaraja *et al.*, 1997; True *et al.*, 1996). De manera formal, ρ se estima a partir de patrones esperados de variación en muestras de secuencias tomadas al azar de una población (Falush *et al.*, 2001; Fearnhead y Donnelly, 2001; Griffiths y Marjoram, 1996; Hey y Wakeley, 1997; Hudson y Kaplan, 1985; Hudson, 1987;2001; Kuhner *et al.*, 2000;2002; McVean *et al.*, 2002; Wakeley, 1997). Sólo se revisarán estos últimos (tabla 3).

Los métodos de estimación de ρ basados en la variación poblacional se pueden dividir en los que buscan patrones usando estadísticos de resumen (Hudson y Kaplan, 1985; Hudson, 1987; Wakeley, 1997), los que utilizan la máxima verosimilitud (ML) o métodos bayesianos completamente (Falush *et al.*, 2001; Griffiths y Marjoram, 1996; Kuhner *et al.*, 2000;2002; Nielsen, 2000; Nielsen sin publicar, descrito en Wall, 2000); y los métodos intermedios, que aprovechan ambas aproximaciones (Fearnhead y Donnelly, 2001; Hey y Wakely, 1997; Hudson, 2001; McVean *et al.*, 2002; Wall, 2000).

A pesar de que los estimadores basados en estadísticos de resumen son fáciles de calcular, ignoran la mayor parte de la información contenida en los datos. Por su parte, aunque las aproximaciones de ML pudieran ser mejores porque aprovechan todo el grupo de datos, aún no es posible utilizar

Tabla 3. Estimadores de la recombinación

Autores	Programa	Distribución / Contacto	Aproximación
Hudson, 1987*	DNAsp	http://www.bio.ub.es/~julio/DnaSP32Inf.html	Estadísticos de resumen
Hudson y Kaplan, 1985a	DNAsp	http://www.bio.ub.es/~julio/DnaSP32Inf.html	Estadísticos de resumen
Wakeley, 1997*	--	http://www.oeb.harvard.edu/faculty/wakeley/	Estadísticos de resumen
Falush *et al.*,2001	--	--	Inf. Bayesiana Completa
Hudson 2001	maxdip y y maxhap	http://home.uchicago.edu/~rhudson1/	Intermedio
Hey y Wakeley, 1997*	SITES	http://lifesci.rutgers.edu/~heylab/index.html	Intermedio
McVean, 2001b	LDHAT	http://www.stats.ox.ac.uk/~mcvean/LDhat/LDhat.html	Intermedio
Fearnhead y Donnelly, 2001	LDHAT-pairwise	http://www.stats.ox.ac.uk/~mcvean/LDhat/LDhat.html	Intermedio
Griffiths y Marjoram, 1996*	recom58	griff@stats.ox.ac.uk	Verosimilitud completa
Kuhner *et al.*, 1999*	RECOMBINE	http://evolution.genetics.washington.edu/lamarc.html	Verosimilitud completa
Nielsen en Wall, 2000*	Baysim	--	Verosimilitud completa

(Continúa)

Tabla 3. Estimadores de la recombinación.

Autores	Programa	Distribución / Contacto	Aproximación
Wall 2000*	hrm	http://home.uchicago.edu/~rhudson1/source/maxdip.html	Verosimilitud completa, Intermedio
Nielsen, 2000	--	http://www.genomics.cornell.edu/faculty/facultybio.cfm?netid=rn28	Verosimilitud e Inf. Bayesiana completas
Nielsen en Wall, 2000	Baysim	--	Verosimilitud e Inf. Bayesiana completas

a. A diferencia de los otros métodos, donde se estima ρ, en éste se estima el parámetro R_m, que es el número de eventos recombinatorios ocurridos en un grupo de secuencias dadas. b. Este método es una extensión de Hudson (2001) en que se implementan diferentes modelos de evolución. * Comparados entre sí con datos simulados en Wall (2000).

estos métodos para muchos datos porque computacionalmente son muy demandantes.

Para estimar sus parámetros casi todos los métodos se basan en el modelo de coalescencia con recombinación (Griffiths y Tavaré, 1994; Hudson, 1983; Nordborg, 2001). En un modelo de coalescencia, la tasa a la cual las muestras o las ramas coalescen en la genealogía depende del tamaño de la muestra (N) y el de la población (N_e). Entre más individuos se muestreen, más pronto se podrá encontrar un par de linajes que coalescan. Después de eso, las mutaciones neutrales pueden colocarse aleatoriamente en ramas de la genealogía a una tasa constante de $\theta / 2$ (para haploides $\theta / 2 = 2N_e\mu$). En ausencia de recombinación y a una tasa de mutación (μ) constante, se espera que las genealogías más grandes tengan un mayor número de mutaciones y que el tiempo al ancestro común más cercano (TMRCA) de todos los alelos muestreados de una población estén relacionados directamente con el nivel de polimorfismo observado en la población.

En presencia de la recombinación, el número de eventos recombinantes que se observa en una población también depende del TMRCA. Mientras la población sea más vieja, mayor es la probabilidad de que los eventos de recombinación hayan ocurrido, ya sea entre individuos o en la historia ancestral de la muestra. Los efectos de la recombinación en una historia muestreada no son una función de la tasa absoluta de recombinación (ρ), sino del producto de la tasa de recombinación por gen por generación (longitud de mapa genético) r, y el tamaño efectivo de la población N_e (Awadalla, 2003). Al hacer uso de la coalescencia, todos los métodos asumen un tamaño constante en la población, independencia de los sitios, evolución neutral y un modelo de sustitución de sitios infinitos, con excepción del método de McVean *et al.* (2002), que puede utilizar diferentes modelos de evolución, incluyendo variación en las tasas de sustitución.

Comparación de los métodos de estimación de ρ

La comparación de los diferentes métodos de estimación de ρ no ha sido tan exhaustiva como entre métodos de detección de la recombinación. Recientemente, Wall (2000) utiliza simulaciones de coalescencia neutra que asume una población panmíctica, grande, sin cambios de tamaño en el tiempo, sin selección y con un modelo de alelos infinitos para comparar los métodos basados en estadísticos de resumen de Hudson (1987) y Wakeley (1997); los métodos intermedios entre verosimilitud y estadísticos

de resumen de Hey y Wakeley (1997), tres diferentes estimadores de Wall (2000) y los métodos de verosimilitud completa de Griffiths y Marjoram (1996), Kuhner *et al.,* (2000) y dos diferentes estimadores de Nielsen (2000; descritos en Wall 2000).

Wall (2000) concluye que el mejor estimador es el de Kuhner *et al.* (2000), porque tiene el segundo sesgo más pequeño, el promedio del error estándar más pequeño y la mayor proporción de estimados correctos medida por el parámetro *g* (Wall, 2000). El método de Nielsen (descrito en Wall, 2000) es también un buen estimador, aunque algunos pocos valores extremos aumentan su error estándar promedio. En contraste, los métodos de Griffiths y Marjoram (1996) y Nielsen (2000) casi siempre subestimaron el valor real (ambos son estimadores que utilizan completamente la verosimilitud).

De los resultados obtenidos por Wall (2000), se puede concluir que independientemente del sesgo y el error promedio estándar, la credibilidad de los estimadores aumenta conforme aumenta ρ, que a bajos valores de ρ algunos son erráticos y que el aumento en la credibilidad es más rápido superando el umbral $\rho = 4\theta$. Sin embargo, en datos reales, esta relación sólo puede darse en N_e muy grandes, con *r* muy pequeñas o con N_e constante y μ muy altas, lo que, si no es incluido en la prueba en cuestión, podría sobreestimar los valores de ρ al confundir μ con *r*. De hecho, Hudson (1987) y Wakeley (1997) previamente notaron que sus estimados de ρ son efectivos sólo en grupos de datos con diversidad y/o recombinación alta ($\rho = \theta \approx 100$).

Por su parte, Fearnhead y Donnelly (2001), al proponer su propio estimado de ρ, basado en verosimilitud completa, hacen una comparación con otros métodos similares basados en la superficie de verosimilitud de éstos y no en su capacidad de estimar ρ a partir datos simulados. De esta manera concluyen que su método es cuatro veces más eficiente que el propuesto por Griffiths y Marjoram (1996) y más preciso que el de Kuhner *et al.* (2000), aunque a decir de Wall (2000) los métodos bayesianos (p. ej. Nielsen, descrito en Wall, 2000) deben probar ser los mejores al no esperar un comportamiento asintótico con respecto a ρ, como los métodos de verosimilitud completa; comportamiento que no es esperado con datos reales.

En cuanto a los métodos que aún no han sido sujetos de comparación, el método de verosimilitud compuesta de Hudson (2001) es promisorio, en tanto que se puede utilizar con grupos de datos grandes (p. ej. Tenaillon *et al.*, 2002). Asimismo la extensión que hacen McVean *et al.,* (2002) al método de verosimilitud compuesta (Hudson, 2001) podría ser aún mejor dado que implementa el uso de diferentes modelos de evolución, abriendo por primera

vez la posibilidad de analizar secuencias que pudieran violar los principios de la coalescencia, es decir casi todas (ver Fearnhead, 2003).

Finalmente, siempre será más sencillo calcular θ que calcular ρ, no sólo porque es conceptualmente más difícil entender los efectos de la recombinación en los patrones de variación observados, sino porque los datos pueden ser consistentes con un amplio rango de posibles tasas de recombinación y valores de θ (Wall, 2000).

Consideraciones finales

La recombinación es un fenómeno omnipresente en los sistemas biológicos. Sin embargo, apenas estamos comenzando a entender sus implicaciones en nuestros diferentes sistemas de estudio. En este sentido, la detección y la medición de la recombinación será de gran ayuda para comprender la dinámica evolutiva de cepas o variedades de bacterias y hongos de la misma especie o especies relacionadas.

El avance en los métodos de detección de la recombinación es notable, las diferentes comparaciones entre ellos (Brown *et al.*, 2001; Maynard-Smith y Smith, 1998; Posada y Crandall, 2001b; Posada, 2002) han permitido un acercamiento más razonado a su aplicación, dependiendo de los datos y los objetivos de estudio. Por ejemplo, al buscar recombinación en linajes viejos putativamente asexuales de rotíferos, hongos y ostrácodos, Gandolfi *et al.* (2003) comparan varios métodos, encontrando coincidencias con los resultados de Posada y Crandall (2001b) y congruencia entre los resultados del programa Reticulate (Jakobsen y Easteal, 1996) y las filogenias reticuladas reconstruidas con el método *split decomposition* (Huson, 1998). En el mismo sentido, tanto *split decomposition* (Huson, 1998; véase el capítulo 3 de este libro) como GENECONV (Sawyer, 1989;1999) coincidieron en la ausencia de recombinación en el *Vibrio cholerae* O139 (Farfan *et al.*, 2002), por lo que el origen de varias clonas patógenas puede no deberse a la recombinación.

A pesar de las limitantes de los métodos de detección de la recombinación, éstos seguirán siendo útiles para todos los casos en los que primero se quiere saber si hay o no recombinación en un grupo de datos y mientras los métodos paramétricos llegan a ser más prácticos. Paralelamente, es de esperarse que los métodos empíricos, basados en caracteres contingentes, como el uso preferencial de codones, cedan lugar a los métodos formales. Por ahora, combinaciones de ambos han arrojado resultados muy interesantes. Por ejemplo, una combinación de métodos filogenéticos de detección de la recombinación

y la comparación de varios genomas de procariontes fotosintéticos sugiere que los genes indispensables para la fotosíntesis se han intercambiado horizontalmente en muchas ocasiones (Raymond *et al.*, 2002).

Por su parte, el desarrollo de métodos para la estimación de la tasa de recombinación aún está en pleno; de hecho, la teoría subyacente sigue desarrollándose (p. ej. Slade, 2001; Wall, 1999; Wiuf y Hein, 1999, 2000; Wiuf, 2002) por lo que es de esperar que surjan nuevas propuestas en el corto plazo. En especial, se espera que las nuevas propuestas analicen la influencia de otras fuerzas evolutivas en la estimación de la recombinación, como la selección natural (Posada *et al.*, 2002) que aún no es integrada a los modelos actuales.

Asimismo, la influencia de la recombinación en la inferencia filogenética está siendo considerada seriamente por varios autores (p. ej. Schierup y Hein, 2000a; Schierup y Hein, 2000b), de tal manera que el problema está generando nuevas propuestas de reconstrucción filogenética que integren la recombinación en sus modelos (p. ej. Huson, 1998; Strimmer *et al.*, 2003).

La importancia de la recombinación también genera preguntas más universales acerca de su origen y permanencia, tanto en las maneras *sui generis* en que ocurre en los procariontes (Redfield, 2001; Vulic *et al.*, 1999), como en el sexo de los eucariontes (Otto y Lenormand, 2002) y la influencia que pudo haber tenido en la evolución de los genomas actuales, de procariontes (Joyce *et al.*, 2002; Ochman *et al.*, 1999; Ochman *et al.*, 2000) y eucariontes (Duret, 2001). La combinación de aproximaciones empíricas y formales, así como el desarrollo de herramientas cada vez más poderosas nos ayudarán a comprender mejor la naturaleza de este proceso y, en consecuencia, a manejar sus resultados de una manera más realista.

Bibliografía

Agapow, P.-M. y A. Burt. 2001. Indices of multilocus linkage disequilibrium. *Molecular Ecology Notes* 1(1-2):101-102.

Akhunov, E.D., A.W. Goodyear, S. Geng, L.L. Qi, B. Echalier, B.S. Gill, Miftahudin, J.P. Gustafson, G. Lazo, S. Chao, O.D. Anderson, A.M. Linkiewicz, J. Dubcovsky, M.L. Rota, M.E. Sorrells, D. Zhang, H.T. Nguyen, V. Kalavacharla, K. Hossain, S.F. Kianian, J. Peng, N.L. Lapitan, J.L. González-Hernández, J.A. Anderson, D.W. Choi, T.J. Close, M. Dilbirligi, K.S. Gill, M.K. Walker-Simmons, C. Steber, P. EJ. McGuire, C.O. Qualset y J. Dvorak. 2003. The organization and rate of evolution of wheat genomes are correlated with recombination rates along chromosome arms. *Genome Research* 13(5): 753-63.

Alokam, S., S.-L. Liu, K. Said y K.E. Sanderson. 2002. Inversions over the terminus region in *Salmonella* and *Escherichia coli*: Is200s as the sites of homologous recombination inverting the chromosome of *Salmonella enterica* serovar typhi. *Journal of Bacteriology* 184(22): 6190-6197.

Anderson, J. y L. Kohn. 1998. Genotyping, gene genealogies and genomics bring fungal population genetics above ground. *Trends in Ecology and Evolution* 13(11): 444-449.

Anisimova, M., R. Nielsen y Z. Yang. 2003. Effect of recombination on the accuracy of the likelihood method for detecting positive selection at amino acid sites. *Genetics* 164(3): 1229-1236.

Ashburner, M. 1989. *Drosophila: A laboratory handbook*. Cold Spring Harbour Laboratory Press, New York.

Awadalla, P. 2003. The evolutionary genomics of pathogen recombination. *Nature Reviews Genetics* 4: 50-60.

Balding, D.J., R.A. Nichols y D.M. Hunt. 1992. Detecting gene conversion: Primate visual pigment genes. *Proceedings of the Royal Society of London Series B* 249(1326): 275-80.

Barker, G.L.A., B.A. Handley, P. Vacharapiyasophon, J.R. Stevens y P.K. Hayes. 2000. Allele-specific pcr shows that genetic exchange occurs among genetically diverse *Nodularia* (cyanobacteria) filaments in the baltic sea. *Microbiology* 146(11): 2865-2875.

Betrán, E., J. Rozas, A. Navarro y A. Barbadilla. 1997. The estimation of the number and length distribution of gene conversion tracts from population DNA sequence data. *Genetics* 146: 89-99.

Boucher, Y., C.L. Nessbø y W.F. Doolittle. 2001. Microbial genomes: Dealing with diversity. *Current Opinion in Microbiology* 4: 285-289.

Bouffard, G.G., J.R. Idol, V.V. Braden, L.M. Iyer, A.F. Cunningham, L.A. Weintraub, J.W. Touchman, R.M. Mohr-Tidwell, D.C. Peluso, R.S. Fulton, M.S. Ueltzen, J. Weissenbach, C.L. Magness y E.D. Green. 1997. A physical map of human chromosome 7: An integrated yac contig map with average sts spacing of 79 kb. *Genome Research* 7: 673-692.

Brown, A.H.D., M.W. Feldman y E. Nevo. 1980. Multilocus structure in natural populations of *Hordeum spontaneum*. *Genetics* 96(2): 523-536.

Brown, C.J., E.C. Garner, A.K. Dunker y P. Joyce. 2001. The power to detect recombination using the coalescent. *Molecular Biology and Evolution* 18(7): 1421-1424.

Burt, A., D.A. Carter, G.L. Koenig, T.J. White y J.W. Taylor. 1996. Molecular markers reveal cryptic sex in the human pathogen *Coccidioides immitis*. *Proceedings of the National Academy of Sciences, USA* 93: 770-773.

Charlesworth, B. y D. Charlesworth. 1979. The evolutionary genetics of sexual systems in flowering plants. *Proceedings of the Royal Society of London Series B* 205: 513-530.

Dobrindt, U. y J. Hacker. 2001. Whole genome plasticity in pathogen bacteria. *Current Opinion in Microbiology* 4: 550-557.

Drouin, G. y G.A. Dover. 1990. Independent gene evolution in the potato actin gene family demonstrated by phylogenetic procedures for resolving gene conversions and the phylogeny of angiosperm actin genes. *Journal of Molecular Evolution* 31(2): 132-150.

Dubnau, D. 1999. DNA uptake in bacteria. *Annual Review of Microbiology* 53: 217-244.

DuBose, R.F., D.E. Dykhuizen y D.L. Hartl. 1988. Genetic exchange among natural isolates of bacteria: Recombination within the phoa gene of *Escherichia coli*. *Proceedings of the National Academy of Sciences, USA* 85(18): 7036-40.

Duret, L. 2001. Why do genes have introns? Recombination might add a new piece to the puzzle. *Trends in Genetics* 17(4): 172-175.

Dykhuizen, D.E., D.S. Polin, J.J. Dun, B. Wilske, V. Preac-Musric, R.J. Dattwyler y B.J. Luft. 1992. *Borrelia bogdorferi* is clonal: Implications for taxonomy and vaccine development. *Proceedings of the National Academy of Sciences, USA* 90: 10162-10167.

Edwards, R.J. y J.F. Brookfield. 2003. Transiently beneficial insertions could maintain mobile DNA sequences in variable environments. *Mol Biol Evol* 20(1): 30-7.

Falush, D., C.Kraft, N.S. Taylor, P. Correa, J.G. Fox, M. Achtman y S. Suerbaum. 2001. Recombination and mutation during long-term gastric colonization by *Helicobacter pylori*: Estimates of clock rates, recombination size, and minimal age. *Proceedings of the National Academy of Sciences, USA* 98(26): 15056-15061.

Farfan, M., D. Minana-Galbis, M.C. Fuste y J.G. Loren. 2002. Allelic diversity and population structure in *Vibrio cholerae* O139 bengal based on nucleotide sequence analysis. *Journal of Bacteriology* 184(5): 1304-1313.

Fearnhead, P. y P. Donnelly. 2001. Estimating recombination rates from population genetic data. *Genetics* 159: 1299-1318.

Fearnhead, P. 2003. Consistency of estimators of the population-scaled recombination rate. *Theoretical Population Biology* 64(1): 67-79.

Feil, E., E.C. Holmes, D.E. Bessen, M.-S. Chan, N.P.J. Day, M.C. Enright, R. Goldstein, D.W. Hood, A. Kalia, C.E. Moore, J. Zhou y B.G. Spratt. 2001. Recombination within natural populations of pathogenic bacteria: Short-term empirical estimates and long-term phylogenetic consequences. *Proceedings of the National Academy of Sciences, USA* 98(1): 182-187.

Feil, E.J., M.C.Enright y B.G. Spratt. 2000a. Estimating the relative contributions of mutation and recombination to clonal diversification: A comparison between *Neisseria meningitidis* and *Streptococcus pneumoniae*. *Research in Microbiology* 151: 465-469.

Feil, E.J., J.M. Smith, M.C. Enright y B.G. Spratt. 2000b. Estimating recombinational parameters in *Streptococcus pneumoniae* from multilocus sequence typing data. *Genetics* 154(4): 1439-1450.

Feil, E.J., J.E. Cooper, H. Grundmann, D.A. Robinson, M.C. Enright, T. Berendt, S.J. Peacock, J.M. Smith, M. Murphy, B.G. Spratt, C.E. Moore y N.P. Day. 2003. How clonal is *Staphylococcus aureus*? *Journal of Bacteriology* 185(11): 3307-16.

Fincham, J.R.S., P.R. Day y A. Radford. 1979. *Fungal genetics*. University of California Press, San Francisco.

Fisher, M.C., G.L. Koenig, T.J. White y J.W. Taylor. 2000. Pathogenic clones versus environmentally driven population increase: Analysis of an epidemic of the human fungal pathogen *Coccidioides immitis*. *Journal of Clinical Microbiology* 38(2): 807-813.

Fisher, R.A. 1932. *The genetical theory of natural selection*. Oxford University Press, Londres.

Fitch, D.H.A. y M. Goodman. 1991. Phylogenetic scanning: A computer assisted algorithm for mapping gene conversions and other recombinational events. *CABIOS* 7: 207-215.

Frandsen, E.V.G., K. Poulsen, M.A. Curtis y M. Kilian. 2001. Evidence of recombination in *Porphyromonas gingivalis* and random distribution of putative virulence markers. *Infection and Immunity* 69(7): 4479-4485.

Gandolfi, A., I.R. Sanders, V. Rossi y P. Menozzi. 2003. Evidence of recombination in putative ancient asexuals. *Molecular Biology and Evolution* 20(5): 754-61.

Gaut, B.S. y A.D. Long. 2003. The lowdown on linkage disequilibrium. *Plant Cell* 15(7): 1502-6.

Geiser, D.M., J.I. Pitt y J.W. Taylor. 1998. Cryptic speciation and recombination in the aflatoxin-producing fungus *Aspergillus flavus*. *Proceedings of the National Academy of Sciences, USA* 95: 388-393.

Gibbs, M.J., J.S. Armstrong y A.J. Gibbs. 2000. Sister-scanning: A monte carlo procedure for assesing signals in recombinant sequences. *Bioinformatics* 16: 573-582.

Gräser, Y., J. Kühnisch y W. Presber. 1999. Molecular markers reveal exclusively clonal reproduction in *Trichophyton rubrum*. *Journal of Clinical Microbiology* 37(11): 3713-3717.

Grassly, N.C. y E.C. Holmes. 1997. A likelihood method for the detection of selection and recombination using nucleotide sequences. *Molecular Biology and Evolution* 14: 239-247.

Griffiths, A.J.F., J.H. Miller, D.T. Suzuki, R.C. Lewontin y W.M. Gelbart. 1999. *An introduction to genetic analysis*. W.H. Freeman and Company, New York.

Griffiths, R.C. y S. Tavaré. 1994. Ancestral inference in population genetics. *Statistical Science* 9: 307-319.

Griffiths, R.C. y P. Marjoram. 1996. Ancestral inference from samples of DNA sequences with recombination. *Journal of Computational Biology* 3: 479-502.

Grigg, M.E. y Y. Suzuki. 2003. Sexual recombination and clonal evolution of virulence in *Toxoplasma*. *Microbes and Infection* 5(7): 685-90.

Guindon, S. y G. Perriere. 2001. Intragenomic base content variation is a potential source of biases when searching for horizontally transferred genes. *Molecular Biology and Evolution* 18(9): 1838-1840.

Halliday, C.L. y D.A. Carter. 2003. Clonal reproduction and limited dispersal in an environmental population of *cryptococcus neoformans* var *gattii* isolates from australia. *Journal of Clinical Microbiology* 41(2): 703-11.

Ham, R.C.H.J.v., J. Kamerbeek, C. Palacios, C. Rausell, F. Abascal, U. Bastolla, J.M. Fernández, L. Jiménez, M. Postigo, F.J. Silva, J. Tamames, E. Viguera, A. Latorre, A. Valencia, F. Morán y A. Moya. 2003. Reductive genome evolution in *Buchnera aphidicola*. *Proceedings of the National Academy of Sciences, USA* 100(2): 581-586.

Haubold, B., M. Travisano, P.B. Rainey y R.R. Hudson. 1998. Detecting linkage disequilibrium in bacterial populations. *Genetics* 150(4): 1341-1348.

Haubold, B. y R.R. Hudson. 2000. Lian 3.0: Detecting linkage disequilibrium in multilocus data. *Bioinformatics* 16(9): 847-849.

Hayes, W.S. y M. Borodovsky. 1998. How to interpret an anonymous bacterial genome: Machine learning approach to gene identification. *Genome Research* 8(11): 1154-71.

Heidelberg, J.F., J.A. Eisen, W.C. Nelson, R.A. Clayton, M.L. Gwinn, R.J. Dodson, D.H. Haft, E.K. Hickey, J.D. Peterson, L. Umayam, S.R. Gill, K.E. Nelson, T.D. Read, H. Tettelin, D. Richardson, M.D. Ermolaeva, J. Vamathevan, S. Bass, H. Qin, I. Dragoi, P. Sellers, L. McDonald, T. Utterback, R.D. Fleishmann, W.C. Nierman y O. White. 2000. DNA sequence of both chromosomes of the cholera pathogen *Vibrio cholerae*. *Nature* 406(6795): 477-83.

Hein, J. 1990. Reconstructing evolution of sequences subject to recombination using parsimony. *Mathematical Bioscience* 98(2): 185-200.

Hein, J. 1993. A heuristic method to reconstruct the history of sequences subject to recombination. *Journal of Molecular Evolution* 36: 396-405.

Heinemann, J.A. y G.F.J. Sprague. 1989. Bacterial conjugative plasmids mobilize DNA transfer between bacteria and yeast. *Nature* 340: 205-209.

Hey, J. y J. Wakeley. 1997. A coalescent estimator of the population recombination rate. *Genetics* 145: 833-846.

Holliday, R. 1964. A mechanism for gene conversion in fungi. *Genetical Research* 5: 282-287.

Holmes, E.C., M. Worobey y A. Rambaut. 1999. Phylogenetic evidence for recombination in dengue virus. *Molecular Biology and Evolution* 16(3): 405-409.

Hudson, R.R. 1983. Properties of a neutral allele model with intragenic recombination. *Theoretical Population Biology* 23: 183-201.

Hudson, R.R. y N.L. Kaplan. 1985. Statistical properties of the number of recombination events in the history of a sample of DNA sequences. *Genetics* 111: 147-164.

Hudson, R.R. 1987. Estimating the recombination parameter of a finite population model without selection. *Genetical Research* 50: 245-250.

Hudson, R.R. 2001. Two-locus sampling distributions and their application. *Genetics* 159: 1805-1817.

Husmeier, D. y F. Wright. 2001. Probabilistic divergence measure for detecting interspecific recombination. *Bioinformatics* 17: S123-S131.

Huson, D.H. 1998. Splitstree: Analyzing and visualizing evolutionary data. *Bioinformatics* 14(1): 68-73.

Imanishi, T. 1996. DNA polymorphisms shared among different loci of the major histocompatibility complex genes. En M. Nei y N. Takahata, (Ed.). *Current issues in molecular evolution*. Institute of Molecular Evolution and Genetics, Pennsylvania State University, Graduated School for Advanced Students, Philadelphia, Hayama.

Jakobsen, I. y S. Easteal. 1996. A program for calculating and displaying compatibility matrices as an aid in determining reticulate evolution in molecular sequences. *Computer Applied to Biosciences* 12: 291-295.

Jakobsen, I.B., S.R. Wilson y S. Easteal. 1997. The partition matrix: Exploring variable phylogenetic signals along nucleotide sequence alignments. *Molecular Biology and Evolution* 14(5): 474-484.

Joyce, E.A., K.Chan, N.R. Salama y S. Falkow. 2002. Redefining bacterial populations: A post-genomic reformation. *Nature Reviews Genetics* 3: 462-473.

Karlin, S. 1998. Global dinucleotide signatures and analysis of genomic heterogeneity. *Current Opinion in Microbiology* 1(5): 598-610.

Kauserud, K. y T. Schumacher. 2001. Outcrossing or inbreeding: DNA markers provide evidence of reproductive mode in *Phellinus nigrolimitatus* (Basidiomycota). *Mycological Research* 105(6): 676-683.

Kimura, M. 1980. A simple method for estimating evolutionary rate of base substitution through comparative studies of nucleotide sequences. *Journal of Molecular Evolution* 16: 111-120.

Kooning, E.V., K.S. Makarova y L. Aravind. 2001. Horizontal gene transfer in prokaryotes. *Annual Review of Microbiology* 55: 709-742.

Koop, B.F., D. Siemieniak, J.M. Slightom, M. Goodman, J. Dunbar, P.C. Wright y E.L. Simons. 1989. Tarsius delta - and beta -globin genes: Conversions, evolution, and systematic implications. *Journal of Biological Chemestry* 264(1): 68-79.

Koski, L.B., R.A. Morton y G.B. Golding. 2001. Codon bias and base composition are poor indicators of horizontally transferred genes. *Molecular Biology and Evolution* 18(3): 404-12.

Koufopanou, V., A. Burt, T. Szaro y J.W. Taylor. 2001. Gene genealogies, cryptic species, and molecular evolution in the human pathogen *Coccidioides immitis* and relatives (Ascomycota, Onygenales). *Molecular Biology and Evolution* 18(7): 1246-1258.

Kroken, S. y J.W. Taylor. 2001. Outcrossing and recombination in the lichenized fungus *Letharia*. *Fungal Genetics and Biology* 34: 83-92.

Kuhner, M.K., J. Yamato y J. Felsenstein. 2000. Maximum likelihood estimation of recombination from population data. *Genetics* 156: 1393-1401.

Kuhner, M.K., J. Yamato y J. Felsenstein. 2002. RECOMBINE v. 1.41. http://evolution.genetics.washington.edu/lamarc/recombine.html.

Lawrence, J.G. y H. Ochman. 1997. Amelioration of bacterial genomes: Rates of change and exchange. *Journal of Molecular Evolution* 44(4): 383-97.

Lawrence, J.G. y H. Ochman. 1998. Molecular archaeology of the *Escherichia coli* genome. *Proceedings of the National Academy of Sciences, USA* 95: 9413-9417.

Lawrence, J.G., R.W. Hendrix y S. Casjens. 2001. Where are the pseudogenes in bacterial genomes? *Trends in Microbiology* 9(11): 535-540.

Lawrence, J.G. y H. Ochman. 2002. Reconciling the many faces of lateral gene transfer. *Trends in Microbiology* 10(1): 1-4.

Ledeerberg, J. y E.L. Tatum. 1946. Gene recombination in *Escherichia coli*. *Nature* 158: 558.

Lewontin, R.C. y K. Kojima. 1960. The evolutionary dynamics of complex polymorphisms. *Evolution* 14: 450-472.

Liu, S.-L. y K.E. Sanderson. 1998. Homologous recombination between rrn operons rearranges the chromosome in host-specialized species of *Salmonella*. *FEMS Microbiology Letters* 164(2): 275-281.

Lorenz, M.G. y J. Sikorski. 2000. The potential for intraspecific horizontal gene exchange by natural genetic transformation: Sexual isolation among genomovars of *Pseudomonas stutzeri*. *Microbiology* 146(12): 3081-3090.

Majewski, J. y F.M. Cohan. 1998. The effect of mismatch repair and heteroduplex formation on sexual isolation in *Bacillus*. *Genetics* 148(1): 13-18.

Marais, G., D. Mouchiroud y L. Duret. 2001. Does recombination improve selection on codon usage? Lessons from nematode and fly complete genomes. *Proceedings of the National Academy of Sciences, USA* 98: 5688-5692.

Marshall, H.D., C. Newton y K. Ritland. 2001. Sequence-repeat polymorphisms exhibit the signature of recombination in lodgepole pine chloroplast DNA. *Molecular Biology and Evolution* 18(11): 2136-2138.

Martin, D. y E. Rybicki. 2000. Rdp: Detection of recombination amongst aligned sequences. *Bioinformatics* 16(6): 562-563.

Maynard-Smith, J. 1992. Analyzing the mosaic structure of genes. *Journal of Molecular Evolution* 34: 126-129.

Maynard-Smith, J., N.H. Smith, M. O'Rourke y B.G. Spratt. 1993. How clonal are bacteria? *Proceedings of the National Academy of Sciences, USA* 90: 4348-43488.

Maynard-Smith, J. y N.H. Smith. 1998. Detecting recombination from gene trees. *Molecular Biology and Evolution* 15(5): 590-599.

Maynard-Smith, J. 1999. The detection and measurement of recombination from sequence data. *Genetics* 153: 1021-1027.

Maynard-Smith, J., E.J. Feil y N.H. Smith. 2000. Population structure and evolutionary dynamics of pathogenic bacteria. *BioEssays* 22: 1115-1122.

McGuire, G., F. Wright y M.J. Prentice. 1997. A graphical method for detecting recombination in phylogenetic data sets. *Molecular Biology and Evolution* 14(11): 1125-1131.

McGuire, G. y F. Wright. 1998. Topal: Recombination detection in DNA and protein sequences. *Bioinformatics* 14(2): 219-220.

McGuire, G. y F. Wright. 2000. Topal 2.0: Improved detection of mosaic sequences within multiple alignments. *Bioinformatics* 16(2): 130-134.

McGuire, G., F. Wright y M.J. Prentice. 2000. A bayesian model for detecting past recombination events in DNA multiple alignments. *Journal of Computational Biology* 7(1-2): 159-170.

McVean, G., P. Awadalla y P. Fearnhead. 2002. A coalescent-based method for detecting and estimating recombination from gene sequences. *Genetics* 160: 1231-1241.

Médigue, C., T. Rouxel, P. Vigier, A. Henaut y A. Danchin. 1991. Evidence for horizontal gene transfer in *Escherichia coli* speciation. *Journal of Molecular Biology* 222: 851-856.

Menotti-Raymond, M., W.T. Starmer y D.T. Sulliva. 1991. Characterization of the structure and evolution of the adh region of *Drosophila hydei*. *Genetics* 127: 355-366.

Milgroom, M.G. 1996. Recombination and the multilocus structure of fungal populations. *Annual Review of Phytopathology* 34: 454-477.

Moszer, I., E.P. Rocha y A. Danchin. 1999. Codon usage and lateral gene transfer in *Bacillus subtilis*. *Current Opinion in Microbiology* 2(5): 524-8.

Mrazek, J. y S. Karlin. 1999. Detecting alien genes in bacterial genomes. *Annals of the NewYork Academy of Sciences* 870: 314-29.

Müller, H.J. 1932. Some genetic aspects of sex. *American Naturalist* 66: 118-138.

Müller, H.J. 1964. The relation of recombination to mutational advance. *Mutation Research* 1: 2-9.

Nagaraja, R., S. Macmillan, J. Kere, C. Jones, S. Griffin, M. Schmatz, J. Terrell, M. Shomaker, C. Jermak, C. Hott, M. Masisi, S. Mumm, A. Srivastava, G. Pilia, T. Featherstone, R. Mazzarella, S. Kesterson, B. McCauley, B. Railey, F. Burough, V. Nowotny, M. D'Urso, D. States, B. Brownstein y D. Schlessinger. 1997. X chromosome map at 75-kb sts resolution, revealing extremes of recombination and gc content. *Genome Research* 7: 210-222.

Nielsen, R. 2000. Estimation of population and recombination rates from single nucleotide polymorphisms. *Genetics* 154: 931-942.

Nordborg, M. 2001. Coalescent theory. En D. J. Balding, M. Bishop y C. Cannings, (Eds.). *Handbook of statistical genetics*. John Wiley and Sons Inc., Chichester, UK. p. 179-212.

Ochman, H., S. Elwyn y N.A. Moran. 1999. Calibrating bacterial evolution. *Proceedings of the National Academy of Sciences, USA* 96(22): 12638-12643.

Ochman, H., J.G. Lawrence y E.A. Groisman. 2000. Lateral gene transfer and the nature of bacterial innovation. *Nature* 405: 299-304.

Ohta, T. y C.J. Basten. 1992. Gene conversion generates hypervariability at the variable regions of kallikreins and their inhibitors. *Molecular Phylogenetics and Evolution* 1: 87-90.

Ohta, T. 2000. Evolution of gene families. *Gene* 259: 45-52.

Otto, S.P. y T. Lenormand. 2002. Resolving the paradox of sex and recombination. *Nature Reviews Genetics* 3: 252-261.

Parkhill, J., B.W. Wren, N.R. Thomson, R.W. Titball, M.T. Holden, M.B. Prentice, M. Sebaihia, K.D. James, C. Churcher, K.L. Mungall, S. Baker, D. Basham, S.D. Bentley, K. Brooks, A.M. Cerdeno-Tarraga, T. Chillingworth, A. Cronin, R.M. Davies, P. Davis, G. Dougan, T. Feltwell, N. Hamlin, S. Holroyd, K. Jagels, A.V. Karlyshev, S. Leather, S. Moule, P.C. Oyston, M. Quail, K. Rutherford, M. Simmonds, J. Skelton, K. Stevens, S. Whitehead y B.G. Barrell. 2001. Genome sequence of *Yersinia pestis*, the causative agent of plague. *Nature* 413(6855): 523-7.

Posada, D. 2001. Unveiling the molecular clock in the presence of recombination. *Molecular Biology and Evolution* 18(10): 1976-1978.

Posada, D. y K.A. Crandall. 2001a. Intraspecific gene genealogies: Trees grafting into networks. *Trends in Ecology and Evolution* 16(1): 37-45.

Posada, D. y K.A. Crandall. 2001b. Evaluation of methods for detecting recombination from DNA sequences: Computer simulations. *Proceedings of the National Academy of Sciences, USA* 98(24): 13757-13762.

Posada, D. 2002. Evaluation of methods for detecting recombination from DNA sequences: Empirical data. *Molecular Biology and Evolution* 19(5): 708-717.

Posada, D. y K.A. Crandall. 2002. The effect of recombination on the accuracy of phylogeny estimation. *Journal of Molecular Evolution* 54: 396-402.

Posada, D., K.A. Crandall y E.C. Holmes. 2002. Recombination in evolutionary genomics. *Annual Review of Genetics* 36: 75-97.

Prado, F., F. Cortes-Ledesma, P. Huertas y A. Aguilera. 2003. Mitotic recombination in *Saccharomyces cerevisiae*. *Current Genetics* 42(4): 185-98.

Przeworski, M. 2002. The signature of positive selection at randomly chosen loci. *Genetics* 160(3): 1179-89.

Ragan, M.A. 2001. On surrogate methods for detecting lateral gene transfer. *FEMS Microbiology Letters* 201(2): 187-191.

Ragan, M.A. 2002. Reconciling the many faces of lateral gene transfer. *Trends in Microbiology* 10(1): 4.

Raymond, J., O. Zhaxybayeva, J.P. Gogarten, S.Y. Gerdes y R.E. Blankenship. 2002. Whole-genome analysis of photosynthetic prokaryotes. *Science* 298(5598): 1616-20.

Redfield, R.J. 2001. Do bacteria have sex? *Nature Reviews Genetetics* 2(8): 634-9.

Rius, N., M.C. Fuste, C. Guasp, J. Lalucat y J.G. Loren. 2001. Clonal population structure of *Pseudomonas stutzeri*, a species with exceptional genetic diversity. *Journal of Bacteriology* 183(2): 736-744.

Robertson, D., P.M. Sharp, F.E. McCutchan y B.H. Hahn. 1995a. Recombination in HIV-1. *Nature* 374: 124-126.

Robertson, D.L., B.H. Hahn y P.M. Sharp. 1995b. Recombination in AIDS viruses. *Journal of Molecular Evolution* 40: 249-259.

Salminen, M.O., J.K. Carr, D.S. Burke y F.E. McCutcham. 1996. Identification of breakpoints in intergenotypic recombinants of HIV type 1 by bootscanning. *AIDS Research and Human Retroviruses* 11(11): 1423-1425.

Sanderson, K.E. y S.L. Liu. 1998. Chromosomal rearrangements in enteric bacteria. *Electrophoresis* 19(4): 569-72.

Satta, Y. 1992. Balancing selection at *hla* loci. En N. Takahata, (Ed.). *Population paleogenetics*. Science Society Press, Tokyo. p. 129-149.

Saunders, N.J., J.F. Peden y E.R. Moxon. 1999. Absence in *Helicobacter pylor*i of an uptake sequence for enhancing uptake of homospecific DNA during transformation. *Microbiology* 145(12): 3523-3528.

Saville, B.J., Y. Kohli y J.B. Anderson. 1998. mtDNA recombination in a natural population. *Proceedings of the National Academy of Sciences, USA* 95: 1331-1335.

Sawyer, S.A. 1989. Statistical tests for detecting gene conversion. *Molecular Biology and Evolution* 6: 526-536.

Sawyer, S.A. 1999. GENECONV: A computer package for the statistical detection of gene conversion. http://www.math.wustl.edu/~sawyer.

Schierup, M.H. y J. Hein. 2000a. Consequences of recombination on traditional phylogenetic analysis. *Genetics* 156: 879-891.

Schierup, M.H. y J. Hein. 2000b. Recombination and the molecular clock. *Molecular Biology and Evolution* 17(10): 1578-1579.

Schierup, M.H., A.M. Mikkelsen y J. Hein. 2001. Recombination, balancing selection and phylogenies in MHC and self-incompatibility genes. *Genetics* 159: 1833-1844.

Schouls, L.M., S. Reulen, B. Duim, J.A. Wagenaar, R.J. Willems, K.E. Dingle, F.M. Colles y J.D. Van Embden. 2003. Comparative genotyping of *Campylobacter jejuni* by amplified fragment length polymorphism, multilocus sequence typing, and short repeat sequencing: Strain diversity, host range, and recombination. *Journal of Clinical Microbiology* 41(1): 15-26.

Siepel, A.C. y B.K. Korber. 1995. Scanning the data base for recombinant HIV-1 genomes. En G. Myers, B. Korber, B. Hahn, K.-T. Jeang y J. Mellors, (Eds.). *Human retroviruses and AIDS 1995: A compilation and analysis of nucleic acid and amino acid sequences*. Theoretical Biology and Biophysics Group, Los Alamos National Laboratory, Los Alamos.

Siepel, A.C., A.L. Macken y B.T.M. Korber. 1995. A computer program designed to screen rapidly for HIV type 1 intersubtype recombinant sequences. *AIDS Research and Human Retroviruses* 11(11): 1413-1416.

Silva, C., L.E. Eguiarte y V. Souza. 1999. Reticulated and epidemic population genetic structure of *Rhizobium etli* biovar *phaseoli* in a traditionally managed locality in Mexico. *Molecular Ecology* 8: 277-287.

Slade, P.F. 2001. Simulation of 'hitch-hiking' genealogies. *Journal of Mathematical Biology* 42: 41-70.

Sneath, P.H.A., M.J. Sackin y R.P. Ambler. 1975. Detecting evolutionary incompatibilities from protein sequences. *Systematic Zoology* 24: 311-322.

Sneath, P.H.A. 1995. The distribution of the random division of a molecular sequence. *Binary* 7: 148-152.

Sneath, P.H.A. 1998. The effect of evenly spaced constant sites on the distribution of the random division of a molecular sequence. *Bioinformatics* 14: 608-616.

Souza, V., T.T. Nguyen, R.R. Hudson, D. Piñero y R.E. Lenski. 1992. Hierarchical analysis of linkage disequilibrium in *Rhizobium* populations: Evidence for sex? *Proceedings of the National Academy of Sciences, USA* 89: 8389-8393.

Stephens, J.C. 1985. Statistical methods of DNA sequence analysis: Detection of intragenic recombination or gene conversion. *Molecular Biology and Evolution* 2: 539-556.

Strimmer, K., K. Forslund, B. Holland y V. Moulton. 2003. A novel exploratory method for visual recombination detection. *Genome Biology* 4(5): R33.

Suerbaum, S., J.M. Smith, K. Bapumia, G. Morelli, N.H. Smith, E. Kuntsmann, I. Dyrek y M. Achtman. 1998. Free recombination within *Helicobacter pylori*. *Proceedings of the National Academy of Sciences, USA* 95: 12619-12624.

Suyama, M. y P. Bork. 2001. Evolution of prokaryotic gene order: Genome rearrangements in closely related species. *Trends in Genetics* 17(1): 10-13.

Takahata, N. 1994. Comments on the detection of reciprocal recombination or gene conversion. *Immunogenetics* 39(2): 146-149.

Tenaillon, M.I., M.C. Sawkins, L.K. Anderson, S.M. Stack, J. Doebley y B.S. Gaut. 2002. Patterns of diversity and recombination along chromosome 1 of maize (*Zea mays* ssp. *mays* L.). *Genetics* 162(3): 1401-1413.

True, J.R., J.M. Mercer y C.C. Laurie. 1996. Differences in crossover frecuency and distribution among three sibling species of *Drosophila*. *Genetics* 142: 507-523.

Valdés, A.M. y D. Piñero. 1992. Phylogenetic estimation of plasmid exchange in bacteria. *Evolution* 46: 641-656.

Vulic, M., R.E. Lenski y M. Radman. 1999. Mutation, recombination and incipent speciation of bacteria in the laboratory. *Proceedings of the National Academy of Sciences, USA* 96: 7348-7351.

Wakeley, J. 1997. Using the variace of pairwise differences to estimate the recombination rate. *Genetical Research* 69: 45-48.

Wall, J.D. 1999. Recombination and the power of statistical tests of neutrality. *Genetical Research* 74: 65-79.

Wall, J.D. 2000. A comparison of estimators of the population recombination rate. *Molecular Biology and Evolution* 17(1): 156-163.

Wall, J.D., P. Andolfatto y M. Przeworski. 2002. Testing models of selection and demography in *Drosophila simulans*. *Genetics* 162(1): 203-16.

Weiller, G.F. 1998. Phylogenetic profiles: A graphical method for detecting genetic recombinations in homologous sequences. *Molecular Biology and Evolution* 15(3): 326-335.

Whittam, T.S., H. Ochman y R.K. Selander. 1983. Multilocus genetic structure in natural populations of *Escherichia coli*. *Proceedings of the National Academy of Sciences, USA* 80(6): 1751-1755.

Wiuf, C. y J. Hein. 1999. The ancestry of a sample of sequences subject to recombination. *Genetics* 151: 1217-1228.

Wiuf, C. y J. Hein. 2000. The coalescent with gene conversion. *Genetics* 155: 451-462.

Wiuf, C., T. Christensen y J. Hein. 2001. A simulation study of the reliability of recombination detection methods. *Molecular Biology and Evolution* 18(10): 1929-1939.

Wiuf, C. 2002. On the minimum number of topologies explaining a sample of DNA sequences. *Theoretical Population Biology* 62: 357-363.

Worobey, M. 2001. A novel approach to detecting and measuring recombination: New insights into evolution in viruses, bacteria, and mitochondria. *Molecular Biology and Evolution* 18(8): 1425-1434.

Zhan, J., R.E. Pettway y B.A. McDonald. 2003. The global genetic structure of the wheat pathogen *Mycosphaerella graminicola* is characterized by high nuclear diversity, low mitochondrial diversity, regular recombination, and gene flow. *Fungal Genetic and Biology* 38(3): 286-97.

Capítulo 10

La especie como unidad evolutiva: uso de marcadores moleculares para su reconocimiento y delimitación, con especial énfasis en microorganismos

René Cerritos Flores

Cuando Darwin publicó en 1859 *El origen de las especies*, se planteó la idea de que las especies pueden surgir de otras a causa de la acumulación de modificaciones por medio de la selección natural, y de otras fuerzas aún no muy claras para el autor. Darwin describió el proceso de especiación pero no propuso alguna forma de reconocimiento y delimitación del producto de tal proceso y, por lo tanto, el libro no tiene un concepto formal de especie.

Es probable que Darwin no haya tenido idea de la gran tarea que nos heredaría: a más de un siglo de la publicación de esta obra aún no se tiene un concepto "natural" que sea funcional para la mayor parte de los organismos, sobre todo aquellos que pertenecen al grupo de los procariontes (Bacteria y Archea), así como algunos grupos de hongos. En este capítulo analizaré los trabajos que se han realizado para reconocer y delimitar especies con base en la teoría evolutiva, haciendo énfasis en la repercusión que han tenido los marcadores moleculares en la resolución de este problema. Asimismo, explicaré cuáles son las principales limitaciones que se presentan al aplicar determinado concepto de especie a organismos con formas de vida muy distintas a las comúnmente estudiadas (plantas y animales), como los microorganismos. Finalmente, haré énfasis en la importancia pragmática que tiene reconocer y delimitar entidades evolutivas, sobre todo en áreas dedicadas a la conservación de la biodiversidad.

De lo natural a lo artificial y de lo universal a lo particular

Desde el punto de vista biológico, una especie se podría describir simplemente como un conjunto de individuos más o menos parecidos que puede variar en tamaño y composición dependiendo del criterio que se use para hacer tal agrupamiento. Cuando un agrupamiento se hace bajo criterios que reconozcan y delimiten el resultado del proceso de especiación estamos ante un grupo natural; en caso contrario, un grupo artificial será aquel en el que hay una inclusión parcial o una inclusión de organismos que pertenecen a más de una entidad evolutiva (Ward, 1998). Estos criterios se fundamentan en teorías o simples patrones que se obtienen en el estudio de modelos biológicos, es decir, de los mismos organismos. Cuando estos modelos logran aplicarse a todas las formas biológicas, estamos ante un criterio universal. De modo contrario, cuando este modelo sólo es aplicable a un grupo muy reducido de formas, se está ante un criterio particular.

Es sorprendente cómo estas agrupaciones cobran importancia cuando se usan criterios naturales y, cuando es posible, universales para reconocerlas y delimitarlas: pensemos que si agrupamos toda la biodiversidad en unidades naturales bajo distintos criterios se pueden resolver muchas de las dudas que en la actualidad siguen rondando en nuestras cabezas. Por ejemplo, si la evolución es gradual o por saltos, si la especiación es simpátrica o alopátrica, cuáles son las velocidades de especiación, cuál es la interacción entre los factores ambientales y la diversidad genética de las poblaciones; éstas son algunas dudas que considero podrían resolverse con sólo aplicar métodos naturales de agrupación.

Criterio 1: semejanza entre los caracteres morfológicos

En el reconocimiento de las especies la forma comúnmente utilizada es intuitivamente muy sencilla: basta con la utilización de cualquier tipo de caracteres morfológicos, conductuales, fisiológicos u otros para descartar o en su defecto integrar individuos a estas entidades biológicas. Por ejemplo, algunas tribus de Nueva Guinea son capaces de reconocer con base en características morfológicas y conductuales cerca de 137 tipos diferentes de aves, que corresponden cada uno a una especie distinta utilizando marcadores moleculares (Avise, 2000). Este tipo de reconocimiento se formalizó antes de la teoría evolutiva,

en especial con Linneo, quien introdujo en 1757 el concepto de *especie* como una forma de reconocer y dar nombre a las distintas entidades biológicas creadas por un ser superior, es decir, un Dios. Este tipo de reconocimiento a través de cualquier carácter morfológico se conoce en honor a Linneo como concepto linneano o morfológico de especie (Mayden, 1997). Posteriormente, Darwin tomó esta misma forma de reconocimiento, aunque seleccionando determinados caracteres como base para sustentar sus explicaciones acerca del origen de las especies: el caso del tamaño y forma del pico de los pinzones de las islas Galápagos es un buen ejemplo para explicar los procesos evolutivos, al menos a esta escala (Darwin, 1859).

Por lo general, este tipo de caracteres se asocian con el método fenético, que tiene como característica principal la de tomar la mayor cantidad posible de caracteres independientemente de su relación ancestro-descendiente. El concepto fenético de especie forma grupos de acuerdo con la similitud de una gran cantidad de caracteres (Sneath, 1976), incluyendo los moleculares.

Criterio 2: los procesos evolutivos originan unidades discretas

Sin duda alguna, el concepto de especie que mejor se apega a la idea de Darwin acerca de la divergencia que se da entre los individuos de una misma población, producto de la selección natural con el paso de las generaciones y que culmina con la aparición de nuevas especies, es el sugerido por Simpson (1961) y Wiley (1978), conocido como **concepto evolutivo** de especie. Éste se define como un linaje (secuencia de poblaciones ancestro-descendiente) que evoluciona separadamente de otros linajes y que tiene una estructura evolutiva propia. Este concepto tiene la facultad de englobar las distintas fuerzas evolutivas que actúan sobre las poblaciones; además, da la pauta para considerar procesos de extinción y sobre todo de especiación.

Sin embargo, por sí mismo no permite reconocer a tales linajes, es decir, este concepto no propone un método capaz de reconocer y delimitar especies (Taylor *et al*., 2000). Así, una conclusión anticipada sería que un concepto de especie que describa y delimite a estos grupos verdaderamente existentes será aquel que englobe la idea de que las especies son unidades evolutivas producto de la especiación con una metodología capaz de reconocerlas y delimitarlas.

Criterio 3: aislamiento reproductivo

Una propuesta que se acerca a la resolución de la problemática conceptual y práctica de las especies es el **concepto biológico** de especie formulado por Dobzhansky (1937) y Mayr (1942). Ellos consideran que la especie es un grupo de poblaciones naturales donde actual o potencialmente existe intercambio genético y que este grupo de poblaciones está aislado reproductivamente de otros grupos análogos. Al detectar eventos de reproducción sexual se infiere que el flujo génico está actuando como fuerza uniformadora en las frecuencias alélicas de las poblaciones y, por lo tanto, se considera que todas estas poblaciones pertenecen a una misma entidad evolutiva. En contraste, cuando se origina una barrera reproductiva o incluso geográfica entre algunas poblaciones, ésta provoca que se diferencien genéticamente hasta que divergen a tal grado que ocurre un proceso de especiación. El método para reconocer especies biológicas es en teoría muy sencillo y basta con registrar la existencia de reproducción sexual, ya sea en el campo o en laboratorio, entre individuos que *a priori* sean considerados de la misma especie.

Criterio 4: un nicho, una especie

El **concepto ecológico** de especie propuesto por Van Valen (1976) se refiere a un linaje (o grupo de linajes fuertemente relacionados) que ocupa una zona adaptativa mínimamente diferente a la de otros linajes. Una especie ecológica puede verse como un grupo de individuos que asimila recursos de una manera semejante y habita en determinadas condiciones abióticas y bióticas. La forma de reconocer las especies ecológicas es correlacionando las condiciones y los recursos del ambiente con algunas características adaptativas, generalmente morfológicas, de cada una de las poblaciones en evaluación. El concepto ecológico se desarrolla dentro de una concepción adaptacionista, en la que se supone que cuando ocurre alguna modificación de las condiciones y/o los recursos de determinado ambiente se crea un nicho vacío, el cual activa la selección natural hasta que la población alcanza un nuevo pico adaptativo asociado a ese nicho y con ello se genera una nueva especie. No obstante, este concepto también puede ser funcional dentro del punto de vista neutral, en cuyo caso se argumenta que los organismos tienen la capacidad de modificar su propio nicho de tal manera que muchos de los caracteres que se observan en los organismos son producto de la deriva génica.

Criterio 5: las relaciones ancestro-descendiente en las poblaciones

En cuanto al **concepto filogenético** de especie, éste ha sido desarrollado por varios autores. Por ejemplo, está el que plantea Cracaft (1983) definido como el grupo más pequeño de organismos que es diagnosticablemente diferente de otros organismos similares y dentro del cual hay un patrón de ancestro-descendiente. Por otra parte, Nixon y Wheler (1990) definen la especie como las agregaciones más pequeñas de poblaciones o linajes diagnosticadas por una única combinación de estados de carácter que reflejen una relación ancestro-descendiente.

Para reconocer y delimitar una especie filogenética primero se identifican y analizan los atributos que reflejen una secuencia ancestro-descendiente en los individuos de los grupos o poblaciones, para así obtener los estados de carácter. Posteriormente se realiza una comparación interpoblacional por medio de una matriz de los perfiles de cada población y finalmente, se divide la matriz en grupos a partir de los diferentes estados de carácter. Cada uno de estos grupos será una especie filogenética. Los resultados de un análisis filogenético se pueden representar en un cladograma donde los grupos o *clusters* más pequeños dentro de esta relación son considerados unidades evolutivas separadas.

El concepto filogenético asume que entre los grupos formados hubo un proceso de especiación, por lo que el flujo génico entre estos grupos es casi nulo y que probablemente la selección natural, la deriva génica y la mutación actúan de manera diferencial sobre los caracteres de los individuos. Los caracteres que se toman para reconocer y delimitar especies filogenéticas pueden ser fenotípicos (morfológicos, fisiológicos, bioquímicos o moleculares) o genotípicos, siempre y cuando reflejen una relación de ancestría en común entre los grupos.

Criterio 6: intercambio genético

El **concepto genético** de especie fue inicialmente propuesto por Ravin (1963) y surgió como una forma de resolver el problema de la aplicabilidad del concepto biológico de especie en microorganismos tales como los procariontes. En este sentido una especie es definida como un grupo de individuos con capacidad de intercambiar información genética por medio de la conjugación, transducción o transformación. Ravin propone que la frecuencia del inter-

cambio genético en procariontes puede ser un índice confiable para conocer las relaciones filogenéticas entre los microorganismos que realizan un proceso de recombinación: donador y receptor (véase el capítulo 9 de este libro). Posteriormente, con el avance de las técnicas en biología molecular, Wayne *et al*. (1987) proponen con base en este concepto una manera de reconocer y delimitar especies: aquellas muestras que tengan más del 70% de hibridización ADN-ADN serán consideradas como una misma especie.

Criterio 7: la cohesividad

El concepto cohesivo de especie propuesto por Templeton (1989, 1998, 1999) es sin duda el trabajo más integrador y completo realizado no sólo en el nivel teórico, sino también en el nivel operacional en el reconocimiento y delimitación de unidades evolutivas. Así, una especie es definida como un linaje (secuencia ancestro-descendiente) donde las fuerzas genéticas y ecológicas crean un grupo cohesivo único de organismos con capacidad de entrecruzamiento. En este sentido podemos redescubrir una especie como un grupo de organismos que tiene una determinada respuesta a las fuerzas ecológicas, es decir, a los factores ambientales (como humedad, temperatura, cantidad de nutrientes y distintos tipos de interacciones) y a las fuerzas genéticas, tales como la mutación, la selección, la deriva o el flujo génico, que interaccionan de tal manera que originan un grupo cohesivo único.

Para demostrar la existencia de especies cohesivas es necesaria la aplicación de bases teóricas y de herramientas moleculares (véase el capítulo 18 de este libro). Una vez que se reconocen los grupos en conflicto en la asignación del estatus de especie cohesiva, se genera una matriz de haplotipos a partir de uno o más marcadores moleculares. Sobre esta matriz de haplotipos se determinan procesos de coalescencia (véase el capítulo 9 de este libro) y a partir de estos puntos se producen anidamientos a los que se asigna el estatus de especie. Finalmente se corrobora la existencia de cohesividad ecológica y genética dentro de los grupos.

El uso de genealogías por haplotipos permite descubrir la variación genética, ya sea dentro de una especie (nivel poblacional) o entre especies. Así, por medio de este método se puede conocer la interfase entre evolución intraespecífica e interespecífica (Templeton, 1993, 2001). Esta interfase es precisamente el límite entre una especie y otra. Para corroborar la cohesividad ecológica se realiza un análisis biogeográfico y filogeográfico (véanse los capítulos 14 y 15 de este libro). Para el primer caso se correlaciona la distribución

espacial de las distintas variantes (en su mayoría morfológicas) de cada uno de los grupos, y para el segundo se correlaciona la distribución espacial de los distintos haplotipos resultantes. Para corroborar la cohesividad genética, son necesarios análisis de diferenciación genética, tales como los estadísticos *F* y sus distintas variantes (Wrigth, 1951; Nei, 1982; Slatkin y Barton, 1989; véase el capítulo 2 de este libro). Estos índices pueden evidenciar flujo génico o, en caso contrario, aislamiento reproductivo entre los grupos en conflicto (Templeton, 1994). Una ventaja que tiene el concepto cohesivo frente a otros conceptos es que los grupos permiten delimitar especies con criterios estadísticos.

Limitaciones de cada método para reconocer y delimitar unidades evolutivas

Cada método que se usa para reconocer y delimitar especies desde el punto de vista evolutivo trata de inferir procesos de especiación, ausencia de flujo génico entre grupos, o presión diferencial de la selección natural y la deriva génica sobre las poblaciones en relación con su ambiente. Sin embargo, en ocasiones estas inferencias para formar grupos naturales resultan no ser fiel reflejo de las unidades evolutivas existentes. A continuación ampliaré cuáles son los problemas más comunes que se presentan en el momento de reconocer y, sobre todo, de delimitar especies.

Problema 1: los caracteres morfológicos y los marcadores moleculares pueden presentar homoplasias

Antes que nada, hay que recordar que en todos los análisis es necesario usar caracteres de tipo homólogo, es decir, que todas las variantes de ese carácter (estados de carácter) provengan de un ancestro común (Li, 1997). Sin embargo, hay ocasiones en que las variantes de determinado carácter provienen de más de un ancestro común, lo que se conoce como homoplasias (véase el capítulo 4 de este libro).

Los caracteres fenotípicos, principalmente morfológicos, conductuales y en algunos casos fisiológicos y bioquímicos, están expuestos a experimentar convergencias evolutivas -un tipo de homoplasia- en donde los caracteres tienden a converger en forma y función independientemente de la relación ancestro-descendiente. Además, un carácter fenotípico es producto no solo de la información genética, también el ambiente y la interacción entre genes y

ambiente lo moldean; por ejemplo, puede ser que las diferencias encontradas a nivel morfológico no existan a nivel genético y lo que se observe sea simplemente producto de una plasticidad fenotípica. Entonces, cuando tomamos caracteres con plasticidad fenotípica o que han experimentado convergencias evolutivas, los límites de una especie pueden ser erróneos. El concepto fenético y morfológico de especie por la metodología que siguen están muy expuestos a producir grupos artificiales.

El reconocimiento de este tipo de caracteres en microorganismos es muy difícil. Las bacterias, arqueas y en algunos casos los hongos presentan una limitada cantidad de caracteres que pueden considerarse como homólogos. En el caso particular de los procariontes, el único método que se aplica en el reconocimiento y delimitación de especies es el fenético, tomando en cuenta caracteres como la forma celular (cocos o bacilos), el tipo de metabolismo (respiración-fermentación), la reacción a las técnicas de tinción (gram+ o gram-), pruebas bioquímicas (lactosa+ o lactosa-), la presencia en determinado ambiente (mesófilas o extremófilas), entre otras. La gran cantidad de homoplasias que pueden existir en estos caracteres hacen de la sistemática bacteriana un sistema completamente artificial, sobre todo en el nivel taxonómico de especie (Cowan, 1968; Priest, 1993; Goodfellow *et al*., 1997). Las especies descritas a través del método fenético enmascaran una gran cantidad de entidades evolutivas. Ward (1998) demostró que la morfoespecie de cianobacteria *Synechococcus lividis* enmascara al menos 11 entidades evolutivas, cuando se aplica el concepto filogenético acompañado de una correlación ambiental (concepto ecológico). A partir de un muestreo a lo largo de un gradiente de temperatura y usando la secuencia 16S rARN para reconstruir la filogenia, llegó a la conclusión de que cada grupo dentro del cladograma corresponde a aislados de un determinado gradiente ambiental.

Los caracteres fenotípicos no son los únicos expuestos a homoplasias, también los marcadores moleculares pueden experimentar convergencias evolutivas, evolución paralela y reversiones a estados ancestrales. Este último tipo de homoplasias ha sido estudiado en secuencias de ADN, en donde es probable que en algunas posiciones (sobre todo en la tercera posición de los codones) existan redundancias por saturación, es decir, si el estado ancestral de una posición en la secuencia es A (adenina), los estados derivados de esa secuencia pueden experimentar subsecuentes cambios (de A a G, de G a C, de C a G y de G a A) hasta llegar nuevamente a presentar la misma base en la misma posición (véase el capítulo 4 de este libro; Li, 1997).

Problema 2: no hay límites estrictos en el aislamiento reproductivo

Respecto al concepto biológico de especie, éste puede tener algunas limitaciones metodológicas, sobre todo cuando se inducen eventos reproductivos de manera artificial, sin contemplar el flujo génico actual. Detectar eventos reproductivos de esta manera no es sinónimo de un intercambio de genes de manera más o menos constante entre los grupos. En estos casos es probable que la reproducción sexual esté desacoplada y permanezca aún cuando la variación genética y el aislamiento geográfico sea muy grande (Taylor *et al*., 2000). Con los análisis de entrecruzamiento se detecta el flujo génico potencial, mas no el flujo génico real y actual (véanse los capítulos 2 y 9 de este libro; Hibbett *et al*., 1995). Un caso que ejemplifica lo anterior es el que se refiere a la especie morfológica del hongo *Pleorotus ostreatus*, en el que se reconocen ocho especies biológicas. Sin embargo, cuando se aplican análisis filogenéticos, usando marcadores moleculares y se correlacionan con las variantes ambientales en cada una de las ocho especies biológicas, se reconocen un total de catorce especies (Vigalis y Sun, 1994).

El concepto biológico pierde completa funcionalidad cuando se aplica a poblaciones en proceso de divergencia y donde en algunos casos puede existir flujo génico que origina grupos híbridos (Rieseberg, 1997). El término híbrido se puede restringir a aquellos organismos que se generan por entrecruzamiento entre individuos de diferentes especies (Rieseberg, 1997; véase el capítulo 13 de este libro). Los híbridos pueden seguir dos rutas antagónicas. Primero, cuando los híbridos son producidos durante un proceso de especiación parapátrica o simpátrica en el que existe entrecruzamiento parcial entre los grupos en divergencia en una zona geográfica y ambientalmente intermedia por largos periodos de tiempo, los híbridos pueden desaparecer debido a la formación de barreras reproductivas definitivas entre los grupos progenitores. Cuando esto sucede se puede hablar de dos entidades evolutivas y un grupo temporal de organismos híbridos (Turner, 1971). Segundo, cuando los híbridos se aíslan reproductivamente de los grupos progenitores se puede originar una nueva especie, siempre y cuando la zona híbrida se mantenga constante en el tiempo. Cuando esto sucede estaríamos ante tres unidades evolutivas: los grupos progenitores y la incipiente especie originada por hibridación. La formación de nuevas especies a través de grupos híbridos se ha documentado extensamente en plantas (Knobloch, 1971; Ellstrand *et al*., 1996; Riesberg, 1997; véase el capítulo 13 de este libro); en el caso de animales y microorganismos se sabe muy poco.

Problema 3: no se puede hablar de un solo factor ambiental que ejerza una presión selectiva en las poblaciones

Los conceptos ecológico y cohesivo pueden no ser infuncionales debido principalmente a la falta de recursos técnicos, estadísticos y conceptuales para analizar conjuntamente todos los factores del hábitat (condiciones y recursos) que puedan estar promoviendo un cambio en las poblaciones, y que finalmente conlleven a un proceso de especiación (Templeton, 2001). Asimismo, aún no se sabe cuál es el efecto que un cambio de una determinada variable ecológica pueda tener en las poblaciones, ni cómo la interacción de distintas variables ambientales afecta y promueve los procesos evolutivos.

Problema 4: un solo marcador molecular puede no ser representativo de la filogenia de la especie

Con respecto al concepto filogenético, el principal problema que se podría presentar al reconocer y, sobre todo, al delimitar unidades evolutivas sería que la filogenia estuviera disociada de la tokogenia (Henning, 1966). El término tokogenia se refiere al flujo génico actual en una población. En tal caso, las especies propuestas a través de la interpretación de un cladograma no corresponderían con los grupos que se formarían si se tomara en cuenta el grado actual de flujo génico. Sin duda el uso de uno o varios caracteres adecuados así como la aplicación correcta de los análisis estadísticos puede evitar esta disociación. David y Nixon (1992) proponen dos análisis filogenéticos para reconocer y delimitar especies. El primero, denominado Análisis de Agregación Poblacional (PAA), intenta separar grupos de acuerdo con atributos que puedan ser únicos en una población determinada (p. ej. fijación de alelos en las poblaciones estudiadas). El segundo, el análisis de Agregación Cladística de Haplotipos (CHA), agrupa las poblaciones que tienen haplotipos idénticos en una sola especie filogenética. En estos análisis se propone el uso de marcadores moleculares y en particular de secuencias de genes para la generación del cladograma. Diversos estudios han puesto a prueba estos dos análisis (De Salle y Vogler, 1994; Escalada *et al*., 1996; Brower, 1999) pero Brower (1999), al examinar las implicaciones que tiene el uso del PAA para delimitar especies a partir de la secuenciación de genes, llega a la conclusión de que este método es inapropiado para inferir las unidades evolutivas. Uno de los problemas que

presenta el PAA es que al usar datos como la secuenciación de genes es casi imposible identificar caracteres discretos (Crother, 1990).

El uso de los marcadores moleculares como herramientas en la reconstrucción filogenética presenta problemas adicionales bajo este concepto de especie. Particularmente, cuando se utilizan secuencias de genes muchas veces se muestrea un solo locus. Cuando este locus presenta una segregación de polimorfismos ancestrales (polimorfismos interespecíficos) la reconstrucción filogenética puede ser incorrecta (Wu, 1991; 1992; Hey, 1994; Templeton, 2001; véase el capítulo 4 de este libro). La segregación de polimorfismos ancestrales hace referencia a los alelos que permanecen en la población aún después de un proceso de especiación. De esta manera, cuando se delimiten los grupos dentro del cladograma, aquellos que compartan este tipo de alelos serán considerados como una sola entidad evolutiva. Para solucionar tal problema se proponen dos opciones: la primera es el análisis de coalescencia (Hudson, 1983,1990; Tajima, 1983; Nei, 1986; Takahata, 1989; Templeton, 1989, 1998, 2001; véase el capítulo 2 de este libro) y la segunda es con el uso de varias genealogías de distintos genes. La concordancia entre las reconstrucciones resultantes en cada uno de estos genes es determinante en la delimitación de especies (Felsenstein, 1985; Wu, 1991, 1992; Avise, 1994, 2000; Hibbett *et al*., 1995; Caccone *et al*., 1996; Giraud *et al*., 1997; O´Donell *et al*; 1998; Streelman *et al*., 1998; Franzot *et al*., 1999; Taylor *et al*., 2000). Usando tres distintas secuencias de genes en el complejo fúngico *Gibberella fujikuroi,* O´Donell *et al*. (1998) encuentran una concordancia entre las tres genealogías generadas, dando como resultado un total de 45 especies, de las que 23 son registradas como nuevas.

Problema 5: bajo el criterio de hibridación ADN-ADN, todos los primates pertenecerÍamos a la misma especie

El concepto genético incurre en dos faltas graves. La primera se refiere a la universalidad del concepto. Si aplicamos el criterio de que dos muestras que presenten más del 70% de hibridación ADN-ADN sean consideradas como la misma especie, todos los géneros de la familia de los primates serían consideradas como miembros de una misma especie (Stanley, 1997). La segunda se refiere a que el criterio de hibridación no está sostenido en algún mecanismo natural que pueda reflejar un proceso de especiación, además, aún no se sabe el papel que juega determinada secuencia en provocar especiación y sobre todo, es probable que esta correlación no sea la misma en diferentes linajes debido a una variación en las tasas de evolución (Maynard Smith, 1995; Em-

bley y Stackebradt, 1997; Goodfellow *et al.*, 1997). Al tomar únicamente la hibridación como criterio para delimitar especies, el concepto genético niega la importancia de determinadas regiones del ADN en conferir una identidad a las especies. Es decir, asume que dentro de ese 70% de ADN total que hibridiza todas las posiciones tienen la misma importancia, independientemente si son secuencias codificantes o no. Sin embargo, es más probable que esa identidad se encuentre en el 30% del ADN total que no hibridiza.

Problema 6: cuando la cohesividad no es tan evidente

Un problema del concepto cohesivo puede surgir cuando las fuerzas ecológicas, las fuerzas genéticas y la reconstrucción genealógica por coalescencia (anidamientos de los haplotipos) delimitan de manera diferente los grupos en evaluación. En los pocos casos en que se ha aplicado el concepto cohesivo jamás se ha visto una discordancia entre los grupos formados a partir de estas tres formas de inferir unidades evolutivas (Gómez-Zurita *et al.*, 2000; Carbone y Kohn, 2001; Templeton, 2001). No obstante, en el caso de que llegara a presentarse un evento de discordancia, habría que volver a replantear los tres tipos de inferencia. Para el caso de la cohesividad ecológica se tendrían que hacer nuevas investigaciones que relacionen ciertos factores ambientales con determinados atributos de los organismos; para el caso de la cohesividad genética, es posible que los marcadores moleculares no sean totalmente neutros o en su defecto, tengan una tasa de evolución diferencial entre los grupos en conflicto, en cuyo caso el uso de marcadores moleculares que sí cumplan estos requisitos es la solución.

Un claro problema que se presenta al usar haplotipos para obtener la genealogía de los grupos es la carencia de una base teórica que iguale estas genealogías generadas por coalescencia con los verdaderos linajes de la población (Templeton, 2001). Asimismo, cuando existe una gran cantidad de alelos en un mismo locus los tiempos de coalescencia pueden ser erróneos (Takahata, 1989; Wu, 1991). Un ejemplo claro es el de la tricotomía de los linajes chimpancé-gorila-humano. Cann *et al.* (1987) encuentran que el tiempo de coalescencia del linaje humano usando ADN mitocondrial es de 200 000 años, mientras que el tiempo de coalescencia usando genes del núcleo es de 400 000 años; el tiempo estimado de divergencia del linaje humano usando otros métodos llega a 5 millones de años (Wu, 1991).

El concepto de especie en microorganismos, un problema aún no resuelto

Los conceptos revisados en este capítulo fueron en su origen desarrollados tomando en cuenta determinadas formas de vida. El concepto biológico, por ejemplo, se construyó tratando de reconocer grupos en organismos con una reproducción sexual, principalmente animales y en algunos casos, plantas (Petersen y Hughes 1999). El concepto ecológico se planteó pensando básicamente en macroorganismos, en los que claramente puede hacerse un análisis de correlación entre la variabilidad morfológica de los grupos con los recursos y condiciones del hábitat (Van Valen, 1976). Es entonces evidente que surge un problema de funcionalidad cuando queremos aplicar un determinado concepto de especie a formas de vida distintas de aquellas para los que inicialmente fue creado.

Los hongos y sus distintas formas de reproducción

En algunos grupos de hongos la aplicación del concepto biológico puede ser impráctico debido a que la formación de meiosporas no se lleva a cabo. Reynolds (1993) afirma que un 20% de toda la diversidad de hongos se reproduce de manera asexual y no produce meiosporas (véase el capítulo 9 de este libro). Algunas otras variedades son homotálicas y pueden producir meiosporas sin la necesidad de una entidad complementaria. En hongos heterotálicos, la presencia de meisporas no es suficiente para inferir entrecruzamiento, por lo que para reconocer y delimitar sus especies, es necesario cultivarlos en el laboratorio para inducir posteriormente su entrecruzamiento. Sin embargo, un problema relevante es que hay variedades que no son cultivables y en algunos casos es casi imposible que en condiciones de laboratorio puedan entrecruzarse.

Considerando esta diversidad reproductiva la solución para reconocer y delimitar especies es el concepto filogenético. En una gran cantidad de estudios, existe concordancia entre las genealogías de distintos genes, de manera que los grupos descritos son considerados entidades evolutivas (Sullivan *et al*., 1995; Giraud *et al*., 1997; Mayden, 1997; O´Donell, 1998; May *et al*., 1999; Fisher *et al*., 2000). Carbone y Kohn (2001), además de usar secuencias de genes, aplican en *Sclerotinia* la teoría de la coalescencia para generar genealogías en cada uno de los loci analizados con la finalidad de evidenciar los tiempos de divergencia tanto de las poblaciones como de las especies del género. Usando secuencias de haplotipos en nueve distintas regiones del ADN nuclear para

385 individuos, obtienen una concordancia entre los nueve loci estudiados, con los mismos tiempos de coalescencia tanto en el nivel poblacional como en el nivel de especie. Asimismo, un análisis filogeográfico de los nueve loci señala que determinados haplotipos son exclusivos de ciertas regiones, perfectamente diferenciadas por una serie de factores ambientales.

Los procariontes se reproducen de manera asexual pero pueden experimentar transferencia horizontal de genes

En los procariontes (Archaea y Bacteria), la aplicación de cada uno de los conceptos de especie se complica más que en cualquier otro grupo de organismos. La forma de vida de la mayoría de los procariontes es completamente diferente de la de los organismos que se han tomado como modelo para desarrollar los conceptos de especie desde el punto de vista evolutivo, a excepción del genético. La diferencia radica en dos procesos: la reproducción y la recombinación. Todos los procariontes sin excepción se reproducen de manera asexual, de tal manera que las células progenitoras son idénticas a las células hijas, a menos que ocurra alguna mutación. Esta reproducción está desligada de la recombinación, de tal manera que el flujo de genes se lleva a cabo de manera lateral u horizontal, ya sea por recombinación homóloga o no homóloga de material genético cromosomal o extracromosomal (plásmidos; ver capítulo 9). Por su parte, la recombinación puede suceder dentro de la misma unidad evolutiva o incluso entre entidades evolutivas filogenéticamente muy distantes (p. ej. entre especies del dominio Bacteria y especies del dominio Eucaria; Bushman, 2001).

La forma actual de reconocer y delimitar las especies procariontes desconoce por completo la dinámica evolutiva, ya que la metodología, completamente fenética, consiste en tomar cualquier tipo de caracteres morfológicos (forma y tamaño), fisiológicos (actividad enzimática) o moleculares (hibridación ADN-ADN o similitud entre las secuencias) para después compararlos con las otras especies fenéticas. Desde hace unos 10 años el único marcador molecular que se utiliza de manera extensiva para tal efecto es el gen que codifica para la subunidad 16s del rARN. Para definir una especie utilizando este marcador se aplica la regla universal del <97%: aquellos aislados que tengan una diferencia mayor a 97% en la secuencia de este gen serán considerados como especies diferentes (Stackebrandt y Goebel, 1994). Esta metodología tiene sus raíces en la hibridación ADN-ADN, propuesta como criterio a partir de que se encontró que la mayoría de las especies definidas con base en análisis

morfofisiológicos presentan una diferencia mínima de 70% cuando se realiza este tipo de hibridación.

El concepto biológico se podría ajustar sin problema a los procariontes, siempre y cuando la frecuencia de recombinación intraespecífica sea muy alta, al grado de generar un monto de variación en las poblaciones equivalente al de la mutación. Es importante mencionar que cuando la contribución de la recombinación a la variación genética es muy baja, la mutación es la fuerza que la está generando, por lo que consideramos que un buen parámetro para definir especies con base el concepto biológico la contribución de la recombinación *vs*. la de la mutación a la variación genética. Se sabe que en especies como *Staphylococcus aureus* -un organismo clonal- la contribución de la mutación a la variación es 15 veces mayor que la de la recombinación (Fiel *et al*., 2003). En cambio, en *Neisseria meningitidis*, *Streptococcus pneumoniae* y *Campilobacter jejuni* la contribución de la recombinación es, respectivamente, 5, 10 y 50 veces más importante que la mutación (Feil *et al*., 2003; Schouls *et al*., 2003). En este caso, la genética de poblaciones puede resolver el problema de delimitación de especies mediante los índices de diferenciación genética y de recombinación (p. ej. desequilibrio de ligamiento) (véase el capítulo 9 de este libro). Por su parte, el concepto filogenético se puede aplicar siempre y cuando la recombinación no homóloga sea muy baja, ya que de lo contrario cada gen tendrá una historia filogenética muy diferente a la de la especie. Lo anterior cobra mayor sentido cuando usamos marcadores moleculares ya que si se utiliza una secuencia genética que tiene la propiedad de transferirse de manera interespecífica, los análisis filogenéticos reflejarán la filogenia de ese gen y no la de la especie.

En la actualidad una de las controversias más fuertes en el estudio evolutivo de los procariontes es la que se refiere a la frecuencia de la recombinación homóloga en poblaciones naturales. Diversos estudios han corroborado que el intercambio genético es extremadamente raro en poblaciones naturales (Robert y Cohan, 1995; Feil *et al*., 1999, 2000; Cohan, 2001, 2002); análisis con diversos marcadores moleculares sugieren que algunos genes o segmentos de genes presentan una frecuencia de recombinación igual o menor que la de la mutación (Selander y Musser, 1990; Maynard Smith *et al*., 1993; Whittam y Ake, 1993; Robert y Cohan, 1995; Feil *et al*., 1999, 2000). Por otro lado, diversos estudios demuestran que la frecuencia de la recombinación es muy alta (Istock *et al*., 1992; Souza *et al*., 1992; Liu *et al*., 1998; Suerbaum *et al*., 1998; Silva *et al*., 1999; Baker *et al*., 2000; Maynard-Smith *et al*., 2000; Frandsen *et al*., 2001; véase la revisión en el capítulo 9 de este libro). Es muy probable que

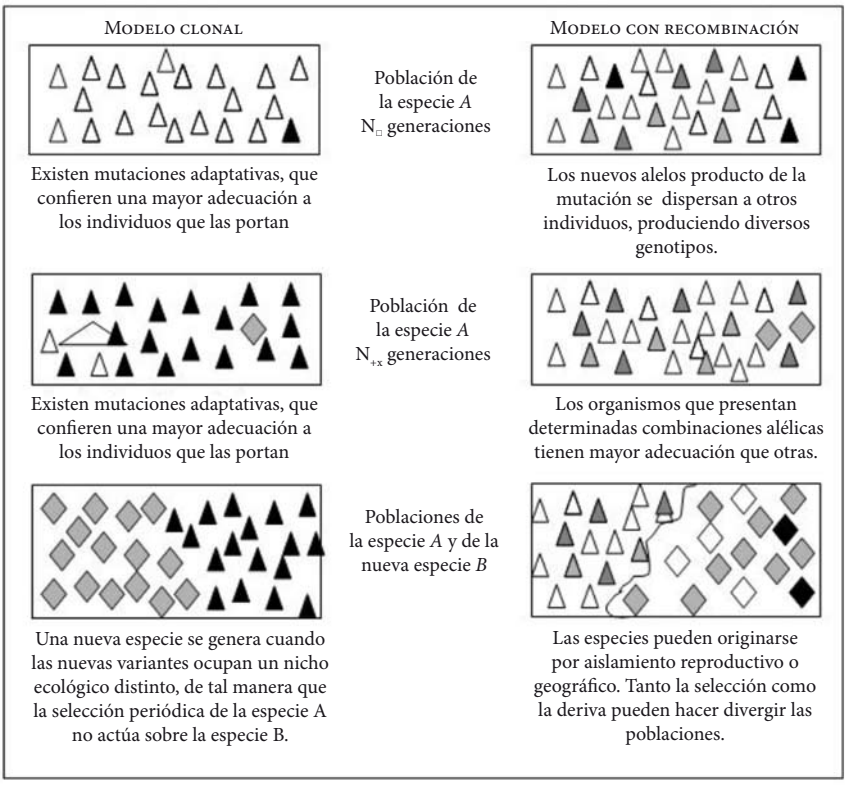
la frecuencia de la recombinación varíe dependiendo de la forma de vida de las entidades evolutivas: en casi todos los ejemplos en los que el intercambio genético es bajo, las especies estudiadas son simbiontes patógenos sumamente virulentos (*Neisseria meningitidis, Streptococcus pneumoniae,* cepas patógenas de *Escherichia coli*), mientras que los grupos donde se encuentra una alta frecuencia de recombinación son simbiontes mutualistas (*Rhizobium etli* y muchas cepas de *Escherichia coli*). Los estudios de recombinación en organismos de vida libre son escasos, reportándose sólo el caso del complejo *Bacillus subtilis*, que demuestra que la recombinación es un evento sumamente frecuente en la naturaleza (Istock *et al*., 1992).

Por otro lado, hay que tener en cuenta que un buen muestreo que abarque la mayoría de las variantes en una población puede ser determinante en los análisis de recombinación. En varios análisis en los que se determinó una baja frecuencia de recombinación, los organismos muestreados se obtuvieron de lugares sumamente distantes y en otros casos se obtuvieron muestras de una sola línea clonal virulenta (Selander y Musser, 1990; Whittam y Ake, 1993; Robert y Cohan, 1995).

En ausencia de recombinación intraespecífica, Cohan (2001, 2002) propone una serie de rutas evolutivas que los procariontes siguen para llegar a procesos de divergencia y consecuentemente a la especiación. Este autor afirma que la mutación es la única fuerza evolutiva capaz de generar variación genética y que la recombinación en cualquier grupo bacteriano es despreciable. El modelo predice que cuando se produce una mutación adaptativa en una población clonal, todos los individuos que no presenten dicha mutación desaparecerán producto de la selección periódica. Debido a la eficiencia que tiene esa nueva variante ante una serie de recursos y condiciones dentro de un espacio dado (nicho ecológico), la selección periódica modifica las frecuencias alélicas drásticamente de 1 (los individuos con la mutación adaptativa) a 0 (todos los individuos que no la presenten). Un proceso de divergencia ocurrirá cuando las mutaciones adaptativas lleven a un subgrupo a modificar la forma de asimilar los recursos y a reaccionar de manera diferencial ante las condiciones del medio. Esta divergencia ocasiona que la selección periódica no afecte al grupo inicial y por lo tanto, estamos ante un proceso de especiación (figura 1). Estos grupos que se forman por una cohesión ecológica son llamados ecotipos; para Cohan pueden ser considerados unidades evolutivas, independientes unas de otras. Sin embargo, en presencia de recombinación intraespecífica este modelo evolutivo propuesto para procariontes puede no ser correcto (figura 1). Las mutaciones adaptativas en conjunción con intercambio genético ocasionan

que la mutación adaptativa se propague en la misma población o incluso en distintas poblaciones que presenten un nicho ecológico distinto, por lo que la frecuencia de los recombinantes en la población estará determinada por procesos de selección natural y deriva génica, más que por selección periódica. Siendo cierto esto, un ecotipo no siempre puede ser considerado como una unidad evolutiva, y puede ser que varios ecotipos conformen una especie.

Figura 1. Dos modelos que explican la dinámica evolutiva de las especies procariontes: el modelo clonal propuesto por Cohan (2000, 2001) y el modelo con recombinación


Modelo clonal
Modelo con recombinación
Población de la especie A N□ generaciones
Existen mutaciones adaptativas, que confieren una mayor adecuación a los individuos que las portan
Los nuevos alelos producto de la mutación se dispersan a otros individuos, produciendo diversos genotipos.
Población de la especie A N+x generaciones
Existen mutaciones adaptativas, que confieren una mayor adecuación a los individuos que las portan
Los organismos que presentan determinadas combinaciones alélicas tienen mayor adecuación que otras.
Poblaciones de la especie A y de la nueva especie B
Una nueva especie se genera cuando las nuevas variantes ocupan un nicho ecológico distinto, de tal manera que la selección periódica de la especie A no actúa sobre la especie B.
Las especies pueden originarse por aislamiento reproductivo o geográfico. Tanto la selección como la deriva pueden hacer divergir las poblaciones.

Consideraciones finales

El reconocimiento y delimitación de especies desde un punto de vista evolutivo es de suma importancia en áreas relativamente nuevas de la biología. La Biología de la Conservación, por ejemplo, busca conocer la riqueza biológica de determinada región en un plano temporal para tomar medidas de protección adecuadas al número de variantes biológicas que ahí existan (véase el capítulo 8 de este libro). La riqueza biológica de una región puede medirse en distintos niveles, desde el nivel genético hasta niveles taxonómicos superiores. En diversos trabajos se plantea la idea de tener unidades formales para medir la biodiversidad (Avise, 1989; Waples, 1991, Moritz, 1994), como las unidades evolutivamente significativas (ESU) que han sido propuestas por varios autores (Ryder, 1986; Waples, 1991; Dizon *et al*., 1992; Avise, 1994; Moritz, 1994; Crandall *et al*., 2000; Fraser y Bernatchez, 2001; véase el capítulo 8 de este libro). Una ESU se describe como un linaje en el que el flujo génico está restringido en relación con otros linajes filogenéticamente muy cercanos, y que tiene una estructura evolutiva propia (Fraser y Bernatchez, 2001). En sentido estricto una ESU es una especie evolutiva y en el inicio del capítulo se planteó la problemática metodológica que tiene el concepto evolutivo de especie para reconocer y delimitar unidades evolutivas. En tal caso la aplicación de los distintos conceptos con una metodología propia pueden conducir al reconocimiento y delimitación de las especies y por consiguiente de las ESUs.

En el presente trabajo se revisó de manera general el uso y aplicación de los distintos conceptos de especie en microorganismos. Sin embargo, grupos tan importantes como los protoctistas o algas no fueron abordados. Asimismo, dentro de los grupos de los hongos y de los dominios Bacteria y Eucaria existe una gran diversidad de rutas evolutivas, por lo que la funcionalidad de los conceptos puede variar dependiendo de ellas.

Un problema que no se abarcó en este trabajo es el que se refiere a la reestructuración de la taxonomía en los niveles inferiores. Buena parte de los grupos que tienen el nivel taxonómico de especie no corresponden a las unidades evolutivas existentes, como el caso de una morfoespecie que puede enmascarar varias unidades evolutivas. Cuando esto ocurre, sin duda hay que tomar decisiones formales y universales para nombrar a éstas y darles un nivel taxonómico determinado.

Los microorganismos han sido agrupados siguiendo un método simplista; el de concebirlos como un mero conjunto de individuos con características

morfológicas y bioquímicas similares. Este método usado por los microbiólogos del siglo XIX no ha cambiado mucho hoy día. El panorama puede ser muy alentador, si al igual que en los macroorganismos vemos estos grupos desde un punto de vista evolutivo, donde cada una de las fuerzas evolutivas interactúan para generar entidades evolutivas con propiedades genéticas, ecológicas y filogenéticas propias. Si asumimos lo anterior cualquier concepto de especie puede ser aplicado en estos grupos para reconocer y delimitar las unidades evolutivas, esas de las que tanto se habla en macroorganismos.

Bibliografía

Avise, C.J. 1989. A role for molecular geneticists in the recognition and conservation of endangered species. *Trends in Ecology and Evolution* 4:279:281.

Avise, C.J. 1994. *Molecular markers, natural history and evolution*. Chapman and Hall, Nueva York.

Avise, C.J. 2000. *Phylogeography. The history and formation of species.* Harvard University Press, Cambridge Massachusetts.

Barker, G. L., B. A. Handley, P. Vacharapiyasophon, J. R. Stevens y P. K. Hayes. 2000. Allele- specific PCR shows that genetic exchange occurs among genetically diverse *Nodularia* (Cyanobacteria) filament in the baltic sea. *Microbiology* 146:2865-75.

Brower, A. V. 1996. Parallel race formation and the evolution of mimicry in *Heliconius* butterflies:a phylogenetic hypothesis from mitochondrial DNA sequences. *Evolution* 50:195-221.

Brower, A. V. 1999. Delimitation of phylogenetic species with DNA sequences:a critique of David and Nixon´s population aggregation analysis. *Systematics Biology* 48 (1):199-213.

Bushman, F. 2001. *Lateral DNA transfer, mechanisms and consequences*. Cold Spring Harbor, Nueva York.

Caccone, A., E. N. Moriyama, J. M. Gleason, L. Nigro y J. R. Powell.1996. A molecular phylogeny for the *Drosophila melanogaster* subgroup and the problem of polymorphism data. *Molecular Biology and Evolution* 13:1224-1232.

Cann, R. L., M. Stoneking y A. C. Wilson.1987. Mitochondial DNA and human evolution. *Nature* 325:31-36.

Carbone, I. y L. M. Kohn. 2001. A microbial population-species interface:nested cladistic and coalescent interface with multilocus data. *Molecular Ecology* 10:947-964.

Cohan, F. M. 2001. Bacterial species and speciation. *Sistematic Biology* 50:513-524.

Cohan, F. M. 2002. What are bacterial species? *Annual Review of Microbiology*. 56:457-487.

Cowan, S. T. 1968. *A dictionary of microbial taxonomy usage*. Oliver and Boyl, Edinburgo.

Cracaft, J.1983. Species concepts and speciation analysis. *Current Ornithology* 1:159-187.

Crandall, K. A., O. R. Binida-Emonds, G. M. Mace y R. K. Wayne. 2000. Considering evolutionary processes in conservation biology. *Trends in Ecology and Evolution* 17:390-395.

Crother, B. I. 1990. Is "some better than none" or do allele frequencies contain phylogenetically useful information? *Cladistics* 6:677-281.

Darwin, C. 1958. *The origin of species*. J. Murray. Londres.

David, J. I. y K. C. Nixon. 1992. Populations, genetics variation, and the delimitation of phylogenetics species. *Systematics Biology* 41:421-435.

De Salle, R. y A. P. Vogler. 1994. Phylogenetics analysis on the edge:the application of cladistic techniques at the population level. En: *Non-neutral evolution:theories and molecular data*. Chapman and Hall, Nueva York.

Dizon, A. E., C. Lockyer, W. F. Perrin *et al*. 1992. Rethinking the stock concept:a phylogeographic approach. *Conservation Biology*, 6:24-36.

Dobzhansky, T.1937. *Genetics and the origin of species*. Columbia University Press, Nueva York.

Ellstrand, N. C., R. Whitkus y L.H.Rieseberg. 1996. Distribution of spontaneous plant hybrids. *Proceedings of the National Academy of Sciences, USA* 93:5090-93.

Embley, T. M. y E. Stackebradt. 1997. Species in practice:exploring uncultured prokaryote diversity in natural samples. En: *Species:the units of biodiversity*. Chapman and Hall, Londres.

Encalada, S. E., P. N. Lahanas, K. A. Bjorndal, A. B. Bolten, M. M. Miyamoto y B. W. Bowen. 1996. Phylogeography and population structure of the Atlantic and Mediterranean green turtle *Chelonia mydas*:a mitochondrial DNA control region sequence assessment. *Molecular Ecology* 5:473-483.

Feil, E. J., J. E. Cooper, H. Grundmann, D. A. Robinson, M.C. Enright, T. Berendt, S. J. Peacock, J. M. Smith, M. Murphy, B. J. Spratt, C. E. Moorey N. P. Day. 2003 How clonal is *Staphylococcus aureus*? *Journal of Bacteriology* 185:3307-16.

Feil, E. J., J. M. Smith, M. C. Enright y B. G. Spratt. 2000. Estimating recombinational parameters in *Streptococcus pneumoniae* from multilocus sequence typing data. *Genetics* 154:1439-1450.

Feil, E. J., M. C. Maiden, M. Achtman y B. G. Sratt. 1999. The relative contributions of recombination and mutation to the divergence of clones of *Neisseria meningitidis*. *Molecular Biology and Evol*ution 16:1469-1502.

Felsenstein, J. 1985. Confidence limits on phylogenies with a molecular clock. *Systematic Zoology* 34:152-161.

Fisher, M. C., G. L. Koening, T. J. White y J. W. Taylor. 2000. A test for concordance between the multilocus genealogies of genes and microsatellites in the patogenic fungus *Coccidiodes immitis*. *Molecular Biology and Evolution* 17:1164-1174.

Frandsen, E. V. , K. Poulsen, M. A. Curtis y M. Kilian, 2001. Evidence of recombination in *Pophyromonas gingivalis* and random distribution of putative virulence markers. *Infection and Immunity* 69:4479-85.

Franzot, S. P., I. F. Salkin y A. Casadevall. 1999. *Cryptococcus neoformans* var. *grubii*: separate varietal status for *Cryptococcus neoformans* serotype A isolates. *Journal of Clinical Microbiology* 37:838-840.

Fraser, D. J., y L. Bernatchez. 2001. Adaptative evolutionary conservation:towards a unified concept for defining conservation units. *Molecular Ecology* 10:2741-2752.

Giraud, T., D. Fortín, C. Lewis, P. Leroux y Y. Brygoo. 1997. RFLP markers show genetic recombination in *Botryoyinia fuckeliana* (*Botrytis cinerea*) and transposable elements reveal two sympatric species. *Molecular Biology and Evolution* 14:1177-1185.

Gómez-Zurita, E. Petitpierre y C. Juan. 2000. Nested cladistic analysis, phylogeography and speciation in the *Timarcha goettingensis* complex (Coleoptera, Chrysomelidae). *Molecular Ecology* 9:557-570.

Goodfellow, M., G. P. Manfio y J. Chun.1997. Towards a practical species concept for cultivable bacteria. En: *Species:the units of biodiversity*. Chapman and Hall, Londres.

Istock, C. A., K. E. Duncan, N. Ferguson y X. Zhou. 1992. Sexuality in a natural population of bacteria - *Bacillus subtilis* challenges the clonal paradigm. *Molecular Ecology* 1:95-103.

Henning, W. 1966. *Phylogenetics systematics*. University of Illinois Press, Urbana.

Hey, J. 1994. Bridging phylogenetics and populations genetics with gene tree models. En: *Molecular ecology and evolution:approaches and applications*. Birkhäuser verlag, Basel.

Hibbett, D.S., Y. Fukumasa-Nakai, A. Tsuneda y M. J. Donoghue. 1995. Phylogenetics diversity in shiitake inferred from nuclear ribosomal DNA sequences. *Mycologia* 87:618-638

Hudson, R. R. 1983. Testing the constant-rate neutral allele model with protein sequence data. *Evolution* 37:203-217.

Hudson, R. R. 1990. Gene genealogies and the coalescent process. *Oxford surveys.* En: *Evolutionary Biology* 7:185-194.

Knobloch, I. W. 1971. Intergeneric hybridisation in flowering plants. *Taxon* 21:97-103.

Li, W.–H. 1997. Molecular Evolution. Sinauer Associates, Massachusetts

Liu, S.-L. y K. E. Sanderson. 1998. Homologous recombination between rrn operons rearranges the chromosome in host-specialized species of *Salmonella. FEMS Microbiology Letters* 164:275-81.

May, G., F. Shaw, H. Badrane y X. Vekemans. 1999. The signature of balancing selection. Fungal mating compatibility gene evolution. *Proceedings of the National Academy of Sciences, USA* 96:9172-9177.

Mayden, R. L. 1997. A hierarchy of species concept:the denouement in the saga of the species problem. En: *Species:the units of biodiversity*. Chapman and Hall, Londres.

Maynard Smith, J., N. H. Smith, M. O´Rourke y B. G. Spratt. 1993. How clonal are bacteria? *Proceedings of the National Academy of Sciences, USA* 90:4384-4388.

Maynard Smith, J.1995. Do bacteria have population genetics? *Society in Genetics Microbiol*ogy 52:1-12.

Maynard-Smith, J., E. J. Feil y N. H. Smith. 200. Population structure and evolutionary dynamics of pathogenic bacteria. *BioEssays* 22:1115-22.

Mayr, E. 1942. *Systematics and the origin of species*. Columbia University Press, Nueva York.

Moritz, C.1994. Defining evolutionary significant units for conservation. *Trends and Ecology and Evolution* 9:373-375.

Nei, M. 1982. Evolution of human races at the gene level. En: *Human Genetics*. Alan R. Liss, Nueva York.

Nei, M. 1986. Stochastic errors in DNA evolution and molecular phylogeny. En: *Evolutionary perspectives and the new genetics*. Alan R. Liss, Nueva York.

Nixon, K.C. y Q. D. Wheler. 1990. An amplification of the phylogenetics species concept. *Cladistics* 6:211-223.

O´Donell; K., E. Cigelnik y H. I. Nirenberg. 1998. Molecular systematics and phylogeography of the *Gibberella fujikuroi* species complex. *Mycologia* 90:465-493.

Petersen, R. H. y K. W. Hughes. 1999. Species and speciation in mushrooms. *BioScience* 49:440-452.

Priest, F. G. 1993. Systematics and ecology of *Bacillus.* En: Bacillus subtilis *and other gram-positive bacteria:biochemistry, physiology and molecular genetics*. American Society for Microbiology, Washington.

Ravin, A. W.1963. Experimental approaches to the study of bacterial phylogeny. *American Naturalist* 97:307-318.

Reynolds, D. R. 1993. The fungal holomorph:an overview. En: *The Fungal holomorph: mitotic, meiotic and pleomorphic speciation in fungal systematics*. Cab International, Londres.

Rieseberg, L. H. 1997. Hybrid origins of plant species. *Annual Review in Ecology and Systematics* 28:359-389.

Robert, M. S. y F. M. Cohan. 1995. recombination and migration rates in natural populations of *Bacillus subtilis* and *Bacillus mojavensis*. *Evolution* 49:1081-1094.

Ryder, O. A. 1986. Species conservation and systematics:the dilemma of subspecies. *Trends in Ecology and Evolution* 1:9-10.

Schouls, L. M., S. Reulen, B. Duim, J. A. Wagenaar, R. J. Willems, K. E. Dingle, F. M. Colles y J. D. Van Embden. 2003. Comparative genotyping of *Campylobacter jejuni* by amplified fragment length polymorphism, multilocus sequence typing, and short repeat sequencing:strain diversity, host range, and recombination. *Journal of Clinical Microbiology* 41:15-26.

Selander, R. K. y J. M. Musser. 1990. Populations genetics of bacterial pathogenesis. En: *Molecular basis of bacterial pathogenesis*. Iglewsky, B. H. Y V. L. Clark, San Diego.

Silva, C., L. E. Eguiarte y V. Souza. 1999. reticulated and epidemic population genetic structure of *Rhizobium etli* biovar phaseoli in a traditionally managed locality in Mexico. *Molecular Ecology* 8:277-287.

Simpson,G.C. 1961. *Principles of animal taxonomy*. Columbia University Press, Nueva York.

Slatkin, M. y N.H. Barton. 1989. A comparison of three indirect methods for estimating average of gene flow. *Evolution* 43:1349-1368.

Sneath, P. H. 1976. Phenetic taxonomy at the species level and above. *Taxon* 25:437-50.

Souza, V. S., T. T. Ngyen, R. R. Hudsen, D. Pinero y R. E. Lensky. 1992. Hierarchical analysis of linkage disequilibrium in *Rhizobium* populations:Evidence for sex? *Proceedings of the National Academy of Sciences, USA* 89:8389-3893.

Stackebrandt, E. y B. M.Goebel. 1994. Taxonomic note:a place for DNA-DNA reassociation and 16s rRNA sequence analysis in the present species definition in bacteriology. *International Journal of Systematics Bacteriology* 44:846-49.

Stanley, J. T.1997. Biodiversity are microbial species threatened? *Current Opinion in Biotechnology* 8:340-345.

Streelman, J. T., R. Zardoya, A. Meyer y S. A. Karl. 1998. Multilocus phylogeny of cichlid fishes (Pisces:Perciformes) evolutionary comparison of microsatellite and single-copy nuclear loci. *Molecular Biolology and Evolution* 15:798-808.

Suerbaum , S., J. M. Smith, K. Bapumia, G. Morelli, N. H. Smith, E. Kuntsmann, I. Dyrek y M. Achtman. 1998:Free recombination within *Helicobacter pylori*. *Proceedings of the National Academy of Sciences, USA* 95:12619-24.

Sullivan, D. J., T. J. Westerneng, K. A. Haynes, D. E. Bennett y D. C. Coleman. 1995. *Candida dubliniensis* sp. nov.:phenotypic and molecular characterization of a novel

species associated with oral candisosis in HIV infected individuals. *Microbiology* 141:1507-1521.

Tajima, F. 1983. Evolutionary relationships of DNA sequences in finite population. *Genetics* 105:437-460.

Takahata, N. 1989. Gene genealogy in three related populations:consistency probability between gene and population trees. *Genetics* 122:957-966.

Taylor, T. A., J. A. Jacobson, S. Kroten, T. Kasuga, D. M. Geiser, D. S. Hibbett, M. C. Fisher. 2000. Phylogenetics species recognition and species concept in fungi. *Fungal Genetics and Biology* 31:21-32.

Templeton, A. R. 1989.The meaning of species and speciation:a genetics perspective. En: *Progress in population genetics and human evolution*. Springer, Nueva york.

Templeton, A. R. 1993. A cladistic analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping. IV. Nested analysis with cladogram uncertainty and recombination. *Genetics* 34:659-669.

Templeton, A. R. 1994. The role of molecular genetics in speciation studies. En: *Molecular ecology and evolution:approaches and applications*. Birkhäuser verlag, Basel, Suiza.

Templeton, A. R. 1998. Species and speciation:geography, populations structure, ecology and gene trees. En: *Endless forms:species and speciation*. Oxford University Press, Oxford.

Templeton, A. R. 1999. Using genes tree to infer species for testable null hypothesis: cohesion species in the *Spalax ehrenbergi* complex. En: *Evolutionary theory and processes:modern perspectives*. Kluwer Academic, Dordrencht.

Templeton, A. R. 2001. Using phylogeographic analyses of gene trees to test species status and processes. *Molecular Ecology* 10:779-791.

Turner, J. R. 1971. Two thousand generations of hybridization in a *Heliconius* butterfly. *Evolution* 25:471-482.

Van Valen, L. 1976. Ecological species, multispecies and oaks. *Taxon* 25:233-239.

Vigalis, R. y B. L. Sun. 1994. Ancient and recent patterns of geographic speciation in the oyster mushroom *Pleorotus* revealed by phylogenetics analysis of ribosomal DNA sequences. *Proceedings of the National Academy of Sciences, USA* 91:4599-4603.

Waples, R. S. 1991. Pacific salmon, *Oncorhynchus* spp. and the definition of species under the endangered species act. *Marine Fisheries Reviews* 53:11-22.

Ward, D. M. 1998. A natural species concept for prokaryotes. *Current Opinion in Microbiology* 1:271-277

Wayne, L. G., D. J.Brenner, R. R. Colwell, P. D. Grimont, O. Kandler, M. I. Krichevsky, L.H. Moore, W. E. Moore, R.G. Murray, E. Stackebrandt. 1987. Report of the ad hoc

committee on reconciliation of approaches to bacteria systematics. *International Journal of Systematics Bacteriology* 37:463-464.

Whittam, T. S. y S. E. Ake. 1993. Genetic polymorphisms and recombination in natural populations of *Escherichia coli*. En: *Molecular paleopopulation biology*. Sci. Soc. Press, Tokio.

Wiley, E.O. 1978. The evolutionary species concept reconsidered. *Systematic Zoology* 27:17-26.

Wrigth, S. 1951. The genetical structure of population. *Annual Eugenics* 15:323-354.

Wu, C. I. 1991. Inferences of species phylogeny in relation to segregation of ancient polymorphisms. *Genetics* 127:429-435.

Wu, C. I. 1992. Gene tress, species trees and the segregation of ancestral alleles. *Genetics* 131:513-513.

Capítulo 11

Ecología evolutiva de bacterias y el concepto de especie: el caso de los rizobios

Claudia Silva y Pablo Vinuesa

En este capítulo se revisan aspectos clave de genética de poblaciones y evolución de procariontes, con énfasis en el grupo de las bacterias fijadoras de nitrógeno conocidas como rizobios. En la primera parte abordamos la ecología y la evolución de la interacción rizobios-leguminosas. La segunda parte toca el intercambio genético en bacterias y los conceptos de especie empleados en biología y en particular en bacteriología. En la tercera parte exponemos los resultados de nuestros estudios sobre genética de poblaciones, filogenia y sistemática molecular de rizobios asociados a leguminosas silvestres y cultivadas; en la última parte presentamos algunas conclusiones a las que hemos llegado sobre la evolución bacteriana a través del estudio de estos organismos.

La interacción rizobios-leguminosas

Ecología de los rizobios

Las bacterias que ocupan los nódulos de plantas leguminosas son un grupo polifilético colectivamente llamado rizobios (Sadowsky y Graham, 1998; Sawada *et al.*, 1993). La característica fenotípica que las distingue de otras bacterias es su habilidad de disparar el desarrollo de órganos fijadores de nitrógeno, llamados nódulos, en las raíces o tallos de plantas hospederas específicas que,

con una excepción, son leguminosas (Long, 1989; van Rhijn y Vyerleyden, 1995; Sadowsky y Graham, 2002).

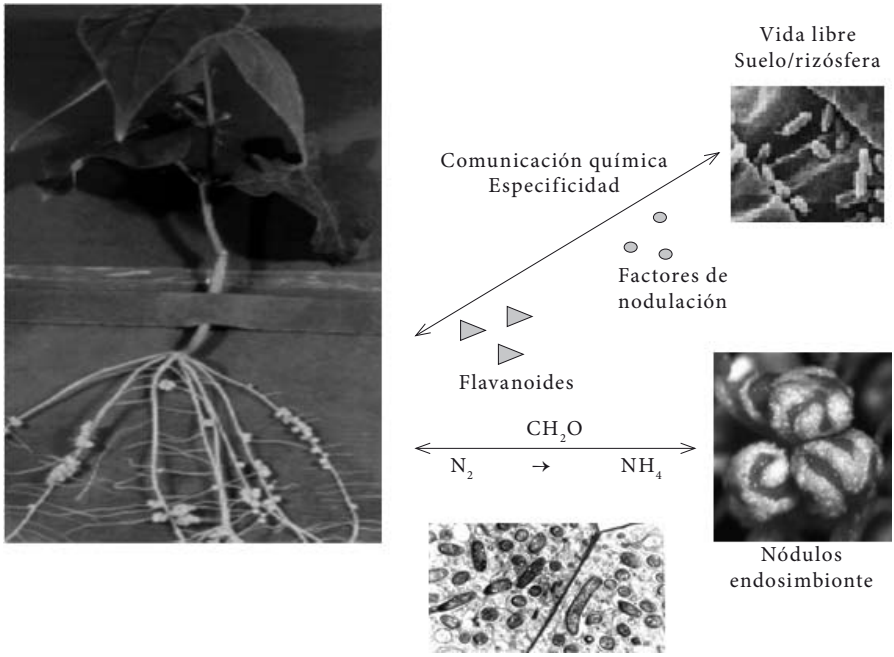
A pesar de que la simbiosis establecida entre rizobios y leguminosas es la interacción planta-bacteria más estudiada (Spaink *et al.*, 1998; Sadowsky y Graham, 2002), el conocimiento que tenemos sobre la ecología de los rizobios es limitado y fragmentado (Sadowsky y Graham, 1998). Los rizobios cubren un espectro ecológico muy amplio: se les encuentra como saprófagos en el suelo y el agua dulce, viven como rizobacterias en la rizósfera de leguminosas y otras plantas, y pueden colonizar el nicho ecológico único de los nódulos en los que, tras ser endocitados por las células del hospedero, se diferencian en bacteroides. En sus historias de vida alterna estadios de vida libre en el suelo y la rizósfera, con estadios endosimbióticos en los que se alojan en un compartimiento intracelular del tejido central del nódulo, denominado simbiosoma (figura 1; Brewin, 1998; Sadowsky y Graham, 1998).

La asociación con los rizobios permite a las leguminosas colonizar exitosamente hábitats deficientes en nitrógeno, donde otras plantas crecen difícilmente. La asociación con las leguminosas beneficia directamente a las poblaciones de rizobios dentro de los nódulos al proporcionarles protección y fuentes de carbono, e indirectamente a las poblaciones de la rizósfera al suministrarles compuestos necesarios para su crecimiento a través de los exudados de las raíces (Long, 1989; Brockwell *et al.*, 1995; Sadowsky y Graham, 2002). Esta interacción se considera una simbiosis mutualista (Long, 1989; Spaink *et al.*, 1998), pero al no ser obligatoria, se trata más de una relación de protocooperación que de mutualismo en sentido estricto (Cheng, 1991).

Ubicación taxonómica de los rizobios y evolución de la simbiosis

En la actualidad la clasificación taxonómica de las bacterias está basada principalmente en las secuencias del gen *rrs* que codifica para la subunidad 16S de rARN (Woese *et al.*, 1990; Weisburg *et al.*, 1991; Ludwig y Schleifer, 1994). Esta clasificación ubica a la mayoría de los rizobios en la subclase α de las Proteobacterias (Willems y Collins, 1993; Young, 1993; Martínez, 1994; Young y Haukka, 1996). En la clasificación actual reconocida en el Manual de Bergey (Madigan *et al.*, 2000) los rizobios se encuentran en cuatro familias del orden Rhizobiales: Rhizobiaceae, que incluye los géneros *Rhizobium* y *Sinorhizobium*; Phyllobacteriaceae, que contiene a *Mesorhizobium*; Hyphomicrobiaceae, que incluye a *Azorhizobium*; y Bradyrhizobiaceae, que incluye

Figura 1. Los rizobios son bacterias heterótrofas aerobias que viven en el suelo como saprófagos. Cuando una leguminosa compatible está presente se establece una estrecha comunicación química que es la base de la simbiosis. Las raíces de las leguminosas exudan flavonoides que activan la expresión de los genes de nodulación de los rizobios, que se encargan de producir y secretar los factores que disparan la organogénesis de los nódulos. Los rizobios infectan las células de los nódulos, dentro de ellas se diferencían en bacteroides y llevan a cabo la fijación biológica de nitrógeno atmosférico. La planta provee de carbohidratos a los bacteroides dentro del nódulo para sostener su metabolismo. Esta interacción es una simbiosis protocooperativa ya que no es obligatoria para ninguno de los dos simbiontes



el género *Bradyrhizobium*. Los géneros de rizobios están filogenéticamente relacionados con otras bacterias de vida libre o patogénicas, como *Rhodopseudomonas*, *Afipia*, *Agrobacterium* y *Brucella* (Willems y Collins, 1993; van Berkum y Eardly, 1998; Sadowsky y Graham, 2002).

Sin embargo, el concepto de rizobios está cambiando, ya que recientemente se reportó una metilobacteria (α-Proteobacteria) noduladora y fijadora, pro-

puesta como *Methylobacterium nodulans* (Sy *et al.*, 2001). Asimismo, se está haciendo patente que la fijación simbiótica de nitrógeno por β-proteobacterias es más común de lo que se había pensado: recientemente se han identificado bacterias de los géneros *Burkholderia* y *Ralstonia* que son capaces de ocupar nódulos de leguminosas y fijar nitrógeno (Chen *et al.*, 2001; Moulin *et al.*, 2001; Chen *et al.*, 2003).

La distribución de los rizobios en las subclases α y β de las Proteobacterias, con raíces filogenéticas profundas y mezclados con bacterias no simbióticas (Young, 1993; Martínez, 1994), genera preguntas sobre la evolución de la simbiosis. ¿Evolucionó independientemente en varios linajes de Proteobacterias? ¿Surgió en su ancestro común y se perdió en sus parientes no simbióticos? ¿Se desarrolló en un linaje y fue transferido a otros? El tiempo de divergencia calculado entre linajes de rizobios, por ejemplo *Bradyrhizobium* y *Rhizobium*, muestra que su ancestro común vivió antes del origen de las plantas terrestres (Turner y Young, 2000), por lo que la hipótesis de un ancestro simbiótico queda prácticamente descartada. Las reconstrucciones filogenéticas de los genes *nod* indican un ancestro común entre ellos (Ueda *et al.*, 1995; Wernegreen y Riley, 1999; Laguerre *et al.*, 2001), por lo que la hipótesis de evolución independiente es poco probable. La hipótesis más apoyada por las evidencias es la de un linaje simbiótico único del que se transfirió lateralmente la información a otros (Young y Johnston, 1989).

Los diferentes compartimientos del genoma de los rizobios

La arquitectura genómica de los individuos de los géneros *Rhizobium* y *Sinorhizobium* está constituida por un cromosoma circular y varios plásmidos (Freiberg *et al.*, 1996; Galibert *et al.*, 2001; González *et al.*, 2003). El cromosoma tiene entre 4 000 y 5 000 kilobases (kb, miles de pares de bases) y los plásmidos pueden ser de tamaño variable, de 25 kb hasta cerca de 2 000 kb. La mayoría de los genes necesarios para establecer la nodulación y fijación de nitrógeno se localizan en uno de los plásmidos, denominado plásmido simbiótico. En *Mesorhizobium* y *Bradyrhizobium* la presencia de plásmidos es menos frecuente; estas bacterias tienen un cromosoma de entre 8 000 y 9 000 kb y la información genética para la simbiosis suele estar agrupada en una región del cromosoma llamada isla simbiótica (Kaneko *et al.*, 2000, 2002; Romero y Broom, 2004). Hasta el momento se han secuenciado los genomas completos de *S. meliloti* 1021, *M. huakuii* MAFF 303099, *B. japonicum* USDA110 (Galibert

et al., 2001; Kaneko *et al.*, 2000, 2002), y están en proceso de anotación los de *R. leguminosarum* 3841 y *R. etli* CFN42. Se conoce la secuencia completa de los plásmidos simbióticos de *S. fredii* NGR234 y *R. etli* CFN42 (Freiberg *et al.*, 1996; González *et al.*, 2003).

Una de las islas simbióticas mejor estudiadas es la de *M. loti* R7A. Esta cepa fue introducida junto con las semillas de *Lotus corniculatus* en un campo de Nueva Zelanda donde no había rizobios nativos capaces de nodular a esta leguminosa. Siete años después se descubrió que los aislados de los nódulos eran genéticamente diversos, pero compartían la misma isla simbiótica cromosomal (Sullivan *et al.*, 1995). El análisis genético de los aislados permitió comprobar que las poblaciones de rizobios nativos, desprovistos de la capacidad de nodular a la leguminosa introducida, adquirieron por transferencia lateral la isla simbiótica de la cepa introducida (Sullivan *et al.*, 1996; Sullivan y Ronson, 1998).

El genoma de los rizobios cuenta con características que le permiten perder y adquirir información simbiótica frecuentemente en la naturaleza, y el descubrimiento de que en condiciones naturales la información simbiótica se transfiere y que la mayoría de los rizobios del suelo carece de información simbiótica, ha marcado un parte-aguas en nuestro entendimiento de la ecología evolutiva de los rizobios (Soberón-Chavez y Nájera, 1988; Segovia *et al.*, 1991, 1995, 1996; Wernegreen y Riley, 1999). El dinamismo del compartimiento simbiótico del genoma de los rizobios les permite adaptarse a condiciones ambientales cambiantes, como la presencia de una nueva leguminosa en un sitio determinado. Los caminos evolutivos de la simbiosis pueden vislumbrarse a través de la comparación de las historias evolutivas de los loci cromosomales de mantenimiento y los loci adaptativos involucrados en la simbiosis (Wernegreen y Riley, 1999).

Genética de poblaciones y evolución en bacterias

El papel del intercambio genético en la evolución bacteriana

Existen dos procesos fundamentales en la transferencia lateral de información entre bacterias. Uno mediado por recombinación homóloga, en el que se sustituyen fragmentos de ADN existentes (véase el capítulo 9 de este libro), y el otro mediado por recombinación sitio específica o aún ilegítima, en la que

se introducen al genoma nuevos fragmentos de ADN (Snyder y Champness, 1997; Gogarten *et al.*, 2002). La frecuencia y la naturaleza de estos procesos tienen consecuencias sobre la estructura genética y los mecanismos de evolución de las bacterias.

En las poblaciones clonales, en las que las células madre dan lugar a dos células hijas por fisión binaria, la variación surge por mutaciones transferidas a los descendientes de las células en que surgieron, y los nuevos linajes se forman por la acumulación de mutaciones sucesivas con el paso de las generaciones. Este tipo de transferencia de información genética se conoce como vertical, mientras que la transferencia horizontal se refiere al movimiento de información genética entre células que no necesariamente comparten un ancestro reciente (Spratt y Maiden, 1999; Gogarten *et al.*, 2002). Se conocen tres mecanismos generales por los que las bacterias pueden transferir lateralmente información genética: conjugación, transformación y transducción (véase el capítulo 9 de este libro). Estos mecanismos sexuales, o más bien parasexuales, difieren en muchos aspectos de los mecanismos sexuales de los eucariontes: son unidireccionales, de un donador a un receptor; se transfiere sólo una fracción del genoma, que puede involucrar desde algunos pares de bases a cientos de kilobases, dependiendo principalmente del mecanismo de transferencia; están desacoplados de la reproducción, por lo que estos eventos no necesariamente ocurren en cada generación: una frecuencia de recombinación de 10^{-5} por gen por generación se consideraría muy alta (Cohan, 1994a; Maynard Smith, 1995; Gogarten *et al.*, 2002).

La contribución relativa de la recombinación en comparación con la de la mutación en la generación de nuevos genotipos varía entre especies y poblaciones bacterianas, y conforme la primera aumenta, la clonalidad de la población disminuye. Existe un amplio espectro de estructuras poblacionales, cuyos extremos son la alta clonalidad, como en *Salmonella enterica*, *Pseudomonas syringae* y *Borrelia borgdorferi*, y la ausencia completa de clonalidad, como en *Neisseria gonorrhoeae*, *N. miningitidis* y *Helicobacter pylori*, pero se considera que la mayoría, incluyendo *Escherichia coli*, *Rhizobium etli, R. gallicum* y varias especies de *Bradyrhizobium*, contiene elementos clonales y no-clonales (Dykhuizen *et al.*, 1992; Souza *et al.*, 1992; Maynard Smith *et al.*, 1993; Gordon *et al.*, 1995; Holmes *et al.*, 1999; Silva *et al.*, 1999; Spratt y Maiden, 1999; Maggi Solcà *et al.*, 2001; Sakar y Guttman, 2004; Vinuesa *et al.*, 2005c).

Si las poblaciones fueran sólo clonales, los diferentes linajes que surgen por mutación serían independientes, estarían sujetos a los efectos de la selección periódica y las filogenias derivadas de sus genes podrían equipararse con la

evolución de los linajes. Su estudio se relacionaría más con el cladismo que con la genética de poblaciones (Maynard Smith, 1995; Levin y Bergstrom, 2000). Si las poblaciones fueran no-clonales, se esperaría una diversidad genética elevada y escaso ligamiento entre genes, los supuestos de la genética de poblaciones encontrarían cabida, pero los análisis filogenéticos de los genes no podrían reconstruir la historia evolutiva de los linajes (Maynard Smith, 1995; Gogarten *et al.*, 2002).

El gran reto en el estudio de la genética de poblaciones y las relaciones evolutivas de bacterias está en determinar las contribuciones relativas de los procesos que generan la variación genética (mutaciones puntuales y recombinación) y las fuerzas que subsecuentemente determinan sus caminos evolutivos (la selección y la deriva génica). Las relaciones entre estas fuerzas evolutivas son complejas, y pueden diferir entre y dentro de especies bacterianas y entre regiones de un genoma (Li, 1997; Nei y Kumar, 2000; Lawrence, 2002).

El estudio de la recombinación en bacterias se ha enfocado en el análisis de la estructura mosaico dentro de los genes, que aporta la evidencia molecular de eventos de recombinación intragénica, y del desequilibrio de ligamiento entre alelos de diferentes loci, que provee de evidencia estadística de recombinación intergénica, así como en la comparación de los árboles genealógicos de diferentes genes que pueden reflejar historias evolutivas distintas para diversas regiones del genoma (Maynard Smith, 1995; Spratt y Maiden, 1999; Feil y Spratt, 2001; véase también el capítulo 9 de este libro). Cuando los datos analizados son secuencias nucleotídicas, en ciertos casos es posible trazar la historia de la recombinación y determinar qué genotipos fueron los donadores y receptores (Dykhuizen y Green, 1991; Yap *et al.*, 1999; Wang y Martinez-Romero, 2000; Brown *et al.*, 2002; van Berkum *et al.*, 2003).

Otro aspecto importante en la evolución de las bacterias es la promiscuidad del intercambio genético, ya que puede ocurrir entre linajes separados por distancias evolutivas mucho mayores que entre los eucariontes (Cohan, 1994a; Feil y Spratt, 2001). La comparación del creciente número de genomas bacterianos secuenciados muestra que han experimentado importantes eventos de transferencia lateral de información, dando origen a cromosomas mosaico de secuencias ancestrales y de genes adquiridos horizontalmente (Nelson *et al.*, 1999; Galibert *et al.*, 2001; PeARN *et al.*, 2001; Wood *et al.*, 2001; Welch *et al.*, 2002). En el caso de las bacterias que cuentan con una parte significativa de su genoma en plásmidos, como es el caso de muchos rizobios, el análisis de las secuencias de plásmidos completos muestra también una estructura de mosaico en la que los elementos móviles como transposones y fagos parecen

tener un papel importante (Freiberg *et al.*, 1996; Freiberg *et al.*, 1997; Galibert *et al.*, 2001; Wood *et al.*, 2001; González *et al.*, 2003).

Se han documentado muchos ejemplos sobre los cambios significativos en el nicho y el fenotipo de una especie provocados por la adquisición de material genético. Por ejemplo, la adquisición de plásmidos, islas de patogenicidad o islas de simbiosis puede provocar el cambio entre ser simbionte comensal, patógeno o mutualista (Sullivan y Ronson, 1998; Wood *et al.*, 2001; Welch *et al.*, 2002). La transferencia de genes ecológicamente adaptativos permite la diversificación y especiación bacteriana, por lo que estos eventos tienen un gran impacto sobre la evolución de las poblaciones de estos organismos. Esta es una de las grandes diferencias y ventajas del intercambio genético en bacterias con respecto al de los eucariontes; las bacterias no tienen que "reinventar la rueda": la transferencia de genes, operones e islas genómicas crean nuevos linajes con combinaciones únicas que pueden explotar nuevos nichos y generar nuevas especies o ecotipos a través de uno o unos pocos eventos evolutivos (Ward, 1998; Lawrence, 1999, 2002; Gogarten *et al.*, 2002).

Este tipo de eventos de recombinación y transferencia lateral son los que pudieron haber ocurrido durante la evolución de la simbiosis entre rizobios y leguminosas, y que ocurren actualmente en el campo. Dentro del grupo de los rizobios existe evidencia de transferencia de islas simbióticas (Sullivan *et al.*, 1995; Sullivan *et al.*, 1996; Vinuesa *et al.*, 2004a) y plásmidos simbióticos entre especies (Segovia *et al.*, 1991; Amarger *et al.*, 1997; Brom *et al.*, 2002), y dentro de especies en condiciones naturales (Silva *et al.*, en preparación; Schofield *et al.*, 1987; Young y Wexler, 1988; Louvrier *et al.*, 1996; Wernegreen *et al.*, 1997; Wernegreen y Riley, 1999; Silva *et al.*, 2003) y de laboratorio (Hooykaas *et al.*, 1977, 1982; Hooykaas y Schilperoort, 1984; Truchet *et al.*, 1984; Martínez *et al.*, 1987; Sivakumaran *et al.*, 1997; Rogel *et al.*, 2001).

Estimaciones de desequilibrio de ligamiento han evidenciado recombinación homóloga intergénica cromosomal en poblaciones locales de *Rhizobium* y *Bradyrhizobium* (Souza *et al.*, 1992; Gordon *et al.*, 1995; Silva *et al.*, 1999, 2003; Vinuesa *et al.*, 2004a) y dentro de *S. meliloti*, *S. medicae* y los tipos A y B de *R. tropici* (Gordon *et al.*, 1995). Hay evidencia de recombinación cromosomal intragénica del gen *glnII* entre *Rhizobium*, *Mesorhizobium* y *Bradyrhizobium* (Turner y Young, 2000), en el gen *rrs* entre especies de *Rhizobium* (Silva *et al.*, en preparación; Eardly *et al.*, 1996; Terefework *et al.*, 1998), en el gen *rrl* de especies de *Bradyrhizobium* (Parker, 2001), así como dentro del operón *rrn* (van Berkum *et al.*, 2003). La evidencia de campo y laboratorio sobre la importancia de la transferencia lateral en la evolución de los rizobios se ha

reforzado con el análisis de los genomas completos y plásmidos simbióticos secuenciados (Freiberg *et al.*, 1997; Kaneko *et al.*, 2000, 2002; Galibert *et al.*, 2001; Wood *et al.*, 2001; González *et al.*, 2003).

A pesar del potencial de intercambio genético en bacterias, los procariontes forman grupos fenotípicos y genéticos discretos que constituyen linajes genéticos coherentes (Cohan, 1994b; Maynard Smith, 1995; Wernegreen y Riley, 1999; véase también el capítulo 10 de este libro). El equilibrio de ligamiento dentro, pero no entre grupos genéticos, sugiere restricciones genéticas a la recombinación.

Aislamiento sexual en bacterias

Se plantea que la transferencia horizontal de información fue la fuerza primaria en la evolución celular (Woese, 2000), lo cual nos lleva a valorarla como una característica inherente a la evolución bacteriana. Inclusive la organización actual de los genes bacterianos en operones (grupos de genes que se transcriben en el mismo mARN, generalmente involucrados en una misma función; Lawrence y Roth, 1996; Snyder y Champness, 1997) sugiere que el genoma de las bacterias ha sido evolutivamente diseñado por y para la transferencia lateral de genes (Lawrence, 2000).

Se ha discutido mucho sobre el papel y la amplitud del aislamiento sexual en bacterias. La visión conservadora de que la transferencia es más frecuente entre organismos relacionados y menos frecuente entre organismos divergentes, es la más aceptada y apoyada por las evidencias, por lo que llamaremos aislamiento sexual a la disminución de la frecuencia del intercambio genético entre especies comparada con la que ocurre dentro de especies (Majewski, 2001). Sin embargo, el grado de aislamiento sexual depende de una multitud de factores que incluyen diferencia en microhábitats, diferencias en sus sistemas de restricción-modificación, divergencia en secuencia y, en el caso de intercambio mediado por plásmidos, fagos o transposones, diferencias entre los rangos de hospedero de los vectores (Cohan, 1996). La transferencia lateral exitosa requiere: 1) disponibilidad de una célula donadora o ADN libre, 2) incorporación del ADN por la célula recipiente, 3) escape del ADN incorporado del sistema de restricción de la célula recipiente, 4) formación del ADN heteroduplex, 5) escape del heteroduplex del sistema de reparación del recipiente, 6) funcionalidad del producto en el entorno genético recipiente (Majewski, 2001). El aislamiento sexual es resultado de la probabilidad de que una o más de estas condiciones no sean satisfechas.

Para la recombinación homóloga, la divergencia entre las secuencias es el principal obstáculo; el grado de aislamiento genético está en función de la divergencia entre ellas (Matic *et al.*, 1996; Vulic *et al.*, 1997). Sin embargo, no hay una discontinuidad obvia en esta función que pueda ser usada para definir un nivel para distinguir las especies (Vulic *et al.*, 1997). La recombinación ilegítima no es afectada por la divergencia entre secuencias, ya que no se basa en el mecanismo de recombinación homóloga, por lo que teóricamente podría ocurrir entre cualquier grupo de organismos. En este caso, las demás barreras al intercambio genético (como proximidad, resistencia a mecanismos de restricción y vectores) parecen ser las más limitantes (Majewski, 2001; Gogarten *et al.*, 2002).

No hay un consenso respecto a cuáles genes son más propensos a intercambio genético y a persistir en los genomas recipientes. Por un lado se propone que los genes más conservados en secuencia e involucrados en el mantenimiento celular, como los rARNs, son potencialmente más propensos al intercambio genético (Cohan, 1994a; Gogarten *et al.*, 2002). Bajo esta visión, la recombinación de genes responsables de las adaptaciones específicas de una población tiene mayor penalidad selectiva que la recombinación de genes que se consideran funcionalmente intercambiables (Cohan, 1994a). Por otro lado, se argumenta que es menos probable que los genes esenciales (como los rARNs) se transfieran exitosamente, ya que el taxón recipiente debe tener ortólogos funcionales que han coevolucionado con el resto de la maquinaria celular y es poco probable que puedan ser desplazados (*hipótesis de la complejidad*; Jain *et al.*, 1999; Woese, 2002)), mientras que los genes bajo selección baja o transitoria (no esenciales) pueden beneficiar a los linajes que los tienen, al otorgarles nuevas capacidades para explotar nuevos nichos (Ward, 1998; Gogarten *et al.*, 1999). El hecho es que, aunque los genes de un genoma pueden tener diferentes probabilidades de ser transferidos exitosamente (Jain *et al.*, 1999; Gogarten *et al.*, 2002; Lawrence, 2002; Woese, 2002), ninguno parece inmune a la transferencia lateral.

Se han reportado casos de transferencia tanto de genes adaptativos, como islas de patogenicidad o simbiosis (Sullivan *et al.*, 1995; Sullivan y Ronson, 1998; Perna *et al.*, 2001; Welch *et al.*, 2002; Vinuesa y Silva, 2004), genes involucrados en adaptaciones a alta temperatura (Nelson *et al.*, 1999) o en resistencia a antibióticos (Dowson *et al.*, 1989; Smith *et al.*, 1992; Salyers *et al.*, 1995), como de genes de mantenimiento indispensables, con funciones metabólicas centrales (Doolittle *et al.*, 1990) o en la maquinaria de síntesis de proteínas (Brochier *et al.*, 2000), en particular en el gen del 16S rARN

(Martínez-Murcia *et al.*, 1992; Mylvaganam y Dennis, 1992; Eardly *et al.*, 1996; Lan y Reeves, 1998; Yap *et al.*, 1999; Sacchi *et al.*, 2002; Hashimoto *et al.*, 2003; Schouls *et al.*, 2003).

El debate sobre el impacto del intercambio genético en la evolución bacteriana continúa, por lo que se plantea que es un tema que se encuentra todavía en su adolescencia conceptual (Lawrence y Hendrickson, 2003) y continúan los esfuerzos por aportar diversas evidencias a favor de posiciones contrastantes. Por ejemplo, para demostrar que la coevolución del rARN con los otros componentes del aparato de traducción no es impedimento para la transferencia horizontal de estos genes, se generó una cepa de *E. coli* con todos sus operones de rARN inactivados, la cual subsiste gracias a la introducción en un vector del operón de rARN de otros géneros bacterianos próximos (*Salmonella typhimurium* o *Proteus vulgaris*), o incluso con algunos dominios provenientes de *Saccharomyces cerevisae* (Asai *et al.*, 1999).

En un análisis basado en genomas completos agrupados por cuartetos, se comparó la topología generada por el rADN 16S con la de los demás genes ortólogos (Daubin *et al.*, 2003). Encontraron que en la mayoría de las comparaciones entre genomas de géneros diferentes, pero relacionados, se converge a una misma topología; sin embargo, al analizar genomas de una misma especie o género la convergencia hacia una misma topología es mucho menor, sugiriendo que a niveles filogenéticos profundos la transferencia lateral no obscurece las relaciones filogenéticas mientras que cuando los organismos están cercanamente relacionados el intercambio genético es más frecuente.

El intercambio genético se reconoce como una importante fuerza de cohesión dentro de las especies y el aislamiento sexual entre ellas puede estar involucrado en los procesos de especiación. A continuación analizamos algunos de los conceptos de especie más significativos.

Generalidades sobre los conceptos de especie

El concepto de especie ha sido uno sobre los que más se ha escrito y uno de los que mayor controversia ha causado en la biología evolutiva (Mayr, 1970; Mayden, 1997; Cohan, 2002; véase también el capítulo 10 de este libro). El término especie se usa tanto para representar una categoría taxonómica como para representar a los individuos que se ordenan en el sistema de clasificación. Estos términos tienen dos categorías filosóficas diferentes. La categoría taxonómica de especie es una clase, que no tiene existencia real y que se usa para ordenar objetos. Las especies como individuos, están limitadas espacial

y temporalmente, tienen cohesión intrínseca, se reproducen, participan en los procesos naturales y cambian con el tiempo.

Gran parte de los problemas sobre los conceptos de especie se relacionan con la confusión de estos dos significados críticamente diferentes (Mayr, 1970; Mayden, 1997; Cohan, 2002). Las reglas de nomenclatura formales han reforzado la visión de las especies como clases y no como individuos. En la literatura taxonómica, sistemática y evolutiva se han desarrollado al menos 22 conceptos de especie que pueden ser agrupados en tres clases básicas: la primera incluye a los conceptos que requieren similitud, la segunda a los que se basan en la monofilia y la tercera a los que requieren aislamiento reproductivo (Hull, 1997; Mayden, 1997). A continuación discutiremos algunos de los más utilizados y representativos. Para una revisión más exhaustiva recomendamos el libro *Species: the units of biodiversity* (Claridge et al., 1997; véase también el capítulo 10 de este libro).

La mayoría de los conceptos de especie están basados en similitud, los más conocidos son los conceptos morfológico, taxonómico, fenético y ecológico. Las especies son, en el concepto morfológico "poblaciones separadas por una discontinuidad en una serie de biotipos", en el concepto taxonómico "todos los especímenes que un taxónomo considere que son miembros de una misma clase", en la definición fenética "el nivel de especie es aquel en el que distintos grupos fenéticos pueden ser observados", y en la definición ecológica "son linajes que ocupan una zona adaptativa mínimamente diferente de la de otros linajes". Estas definiciones se basan en determinar que la variación de un conjunto de caracteres es menor dentro de un grupo que entre grupos, y las especies son los grupos más pequeños que exhiben el grado apropiado de similitud.

Su principal problema es que no dan razones para escoger un nivel de similitud sobre otro. Además, estas definiciones no permiten estudiar las especies como entidades históricas que forman linajes; el tiempo simplemente no entra en la ecuación. Si tomamos en cuenta que la evolución es un proceso que ocurre a lo largo del tiempo, ninguna de estas definiciones puede tratar a las especies como resultado de la evolución. En gran medida se mantiene una visión fijista y esencialista pre-evolutiva (Hull, 1997; Mayden, 1997). Sin embargo, estos conceptos son aplicables a todos los organismos, por lo que han sido los más utilizados por los taxónomos y biólogos en general.

Entre los conceptos de especie basados en monofilia, el más representativo es el concepto filogenético, en el que una especie es el grupo monofilético más inclusivo definible por al menos una autoapomorfía (carácter derivado

compartido). Las especies se delimitan por la distribución de caracteres diagnósticos fijos y se caracteriza por definir grupos monofiléticos. Este concepto es universalmente aplicable a todos los organismos, aunque queda la duda de si a todos los grupos monofiléticos más inclusivos se les quiere aplicar la categoría de especie. Además de la falta de un rango en el cual definir un grupo monofilético como especie, este concepto no tiene dimensión temporal, por lo que su significancia teórica es limitada (Hull, 1997; Mayden, 1997). Los conceptos basados en monofilia surgen de la creciente rama de la sistemática filogenética y de la necesidad de una definición de especie basada en linajes, operacional y libre de procesos.

El concepto biológico de especie es el más representativo de los conceptos basados en el aislamiento reproductivo; plantea que las especies son grupos de poblaciones que comparten un mismo acervo genético y que están aisladas reproductivamente de las otras especies (Mayr, 1970). La elaboración de este concepto fue fundamental en el pensamiento biológico, ya que se sustenta en la teoría evolutiva. En primer lugar, reemplaza la interpretación de las especies como clases de objetos por la de individuos con realidad y cohesión interna debida a la evolución histórica de un programa genético compartido por los miembros de la especie. En segundo lugar, provee de un criterio operacional para definir a las especies, el flujo génico y su interrupción. Sin embargo, este concepto tiene muchas desventajas para ser utilizado como concepto de especie. Muchas han sido revisadas exhaustivamente (Hull, 1997; Mayden, 1997; Cohan, 2002), por lo que sólo mencionaremos que no es universalmente aplicable -todos los organismos con reproducción asexual quedan excluidos-, es un concepto relativista -una especie existe en función de otra- y con poca perspectiva de linaje (Hull, 1997; Mayden, 1997).

Un concepto de especie que tiene gran importancia teórica y que no cae en ninguna de las categorías antes descritas es el concepto evolutivo. Bajo este concepto, una especie es un linaje que evoluciona separadamente de otros, que tiene sus propias tendencias y papeles evolutivos. Esta definición de especie es universalmente aplicable, tiene perspectiva de linaje y se fundamenta en la teoría de la evolución. El principal problema que presenta es que es operacionalmente difícil, ya que se plantea que es la evolución la que produce linajes con identidad y cohesión, y este proceso es difícil de medir (Hull, 1997; Mayden, 1997). Para introducir mecanismos operacionales en el concepto evolutivo, se creó el concepto cohesivo de especie, en el que una especie es un linaje evolutivo cuyos límites surgen por fuerzas genéticas y ecológicas que crean comunidades reproductivas cohesivas. Bajo esta definición,

la tarea básica del investigador es identificar los mecanismos de cohesión que mantienen a un grupo como un linaje evolutivo (Templeton, 1989). Sin embargo, las definiciones científicas teóricas no suelen introducir los criterios de aplicación en sus definiciones; incluir mucha terminología o datos empíricos dentro del cuerpo de una definición puede ser contraproducente (Hull, 1997), por lo que consideramos que el concepto evolutivo provee de una definición evolutiva más clara.

Dada la necesidad teórica y práctica de contar con un cuerpo teórico que nos permita manejar la diversidad biológica, es necesario establecer un concepto de especie que pueda ser aplicado a la totalidad del mundo vivo. La solución no es agrupar distintos componentes de varios conceptos. Por el contrario, el agrupar criterios discordantes en un solo concepto ha generado en gran medida el debate alrededor del concepto de especie (Mayden, 1997). Consideramos que el concepto evolutivo es mejor, ya que tiene fundamento teórico y es aplicable a todo el mundo vivo. Es el único que es consistente con el conocimiento teórico y empírico que tenemos sobre la diversificación del mundo vivo, es consistente con el estatus de individuo de las especies y engloba todos los tipos de entidades biológicas. El principal inconveniente es que es poco operacional en su definición, por lo que estamos de acuerdo con Mayden (1997) en que el concepto evolutivo sea utilizado como el concepto de especie primario y que los demás sean aplicados como conceptos secundarios supeditados al concepto primario (figura 2).

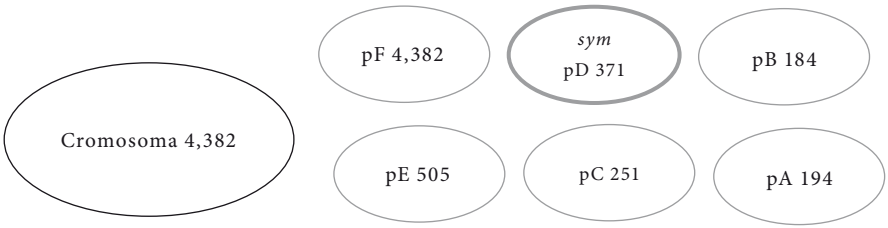
Los conceptos de especie en bacterias

El desarrollo de técnicas cada vez más avanzadas para explorar la diversidad microbiana nos ha permitido conocer con mucha mayor profundidad la biología de las bacterias. Estas herramientas han aportado luz sobre la complicada manera en que este enorme grupo de organismos se organiza, desde el nivel ecológico hasta el molecular. El desarrollo de técnicas que permiten conocer la diversidad bacteriana no cultivable y la observación *in situ* de sus comunidades, ha abierto nuevos horizontes en la ecología microbiana.

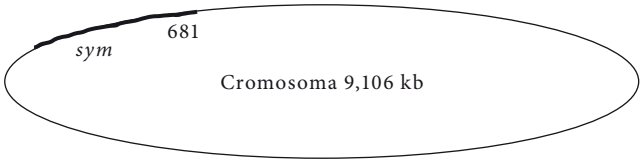
Existen más de 100 genomas bacterianos completamente secuenciados (www.ncbi.nlm.nih.gov). Su análisis comparativo ha mostrado la gran plasticidad genómica y genética de las bacterias, también ha mostrado que el genoma bacteriano es un mosaico y que la transferencia lateral de información genética juega un papel importante en su evolución (Lawrence, 1999; Doolittle, 2002). Una conclusión derivada del mosaicismo presente en los genes y genomas de

Figura 2. A) Genoma de la cepa tipo de *Rhizobium etli* CFN42 aislada de un nódulo de frijol. El genoma consiste de un cromosoma circular y seis plásmidos (pA-F). La mayor parte de los genes involucrados en la simbiosis se localizan en uno de los plásmidos, llamado plásmido simbiótico (*sym*). B) Genoma de la cepa *Bradyrhizobium japonicum* USDA 110 aislada de un nódulo de soya. El genoma consiste de un solo cromosoma circular y los genes simbióticos se encuentran agrupados en una región llamada isla simbiótica (*sym*)

A) Genoma de *Rhizobium etli* CFN42^{T} (kb)



B) Genoma de *Bradyrhizobium japonicum* USDA 110 (kb)



las bacterias es que las filogenias derivadas de un solo gen no necesariamente reflejan la historia evolutiva de los organismos (Nelson *et al.*, 1999; Doolittle, 2002). Esta conclusión tiene fuertes repercusiones sobre la sistemática bacteriana, ya que ésta se basa en gran medida en las relaciones filogenéticas de un solo gen, el gen *rrs*.

Aún con el gran avance que han experimentado recientemente las áreas de la ecología y evolución bacteriana, el concepto de especie empleado en bacteriología se ha mantenido al margen de muchos de estos descubrimientos (Cohan, 2002). La sistemática bacteriana no ha incorporado los conceptos de especie basados en teorías evolutivas. El Comité Internacional de Sistemática Bacteriana es el encargado de establecer y recomendar los estándares mínimos para la descripción válida de especies bacterianas y pretende proveer a los bacteriólogos de una definición uniforme de especie procarionte (Stackebryt

et al., 2002). En la última nota taxonómica publicada por el comité, las especies se definen de manera pragmática, operacional y universalmente aplicable: una especie es una categoría que circunscribe un grupo (preferentemente) genómicamente coherente de aislados individuales que comparten un alto grado de similitud en (varias) características independientes, probadas bajo condiciones altamente estandarizadas. La similitud ADN:ADN se mantiene como el parámetro reconocido para delimitar especies. Se ha adoptado un nivel de 70% de similitud como el estándar reconocido (*gold standard*) para determinar si dos cepas pertenecen a la misma especie y se ha utilizado el 3% divergencia en secuencia del gen *rrs* como punto de corte para separar especies (Stackebryt y Goebel, 1994; Stackebryt *et al.*, 2002). Así en bacteriología se ha usado un concepto de especie arbitrario (Maynard Smith, 1995; Ward, 1998). Es un concepto generado *ad hoc* para bacterias y no es aplicable a todos los seres vivos, si se aplicara este concepto de especie a otros seres vivos encontraríamos que, por ejemplo, todos los primates pertenecemos a la misma especie (Maynard Smith, 1995; Ward, 1998).

La definición de especie bacteriana actual se basa en el diagnóstico de similitud entre cepas. La concepción tipológica de las especies es persistente en bacteriología, ya que las especies se siguen considerando entidades estáticas (que no evolucionan) definidas por atributos típicos o esenciales (Mayr, 1970; Ward, 1998). Se designan cepas tipo y sus caracteres fenotípicos y genotípicos se consideran típicos de todos los organismos que pertenecen a esa especie. El concepto tipológico ignora el hecho de que las especies están compuestas de poblaciones naturales con una organización interna que les provee de una estructura que va más allá de una mera agregación de individuos (Mayr, 1970). El reemplazo del pensamiento tipológico por el pensamiento poblacional ha sido una de las mayores revoluciones conceptuales en biología (Mayr, 1970), pero al parecer ésta no ha llegado a la sistemática bacteriana.

En la definición de especie actual no se incluyen las nuevas evidencias y visiones sobre la evolución bacteriana, pero algunos biólogos evolutivos han tratado de proponer conceptos de especie bacteriana basados en teorías evolutivas y ecológicas. Dykhuizen y Green (1991) propusieron que se puede aplicar el concepto biológico de especie a las bacterias. Sugieren que las bacterias se pueden delimitar como grupos de cepas que recombinan entre sí pero no con cepas de otros grupos y proponen una aproximación filogenética, basada en secuencias de varios loci, para identificar los grupos que han, o no han, intercambiado genes. Esta propuesta tiene la ventaja de incorporar la teoría evolutiva a la sistemática bacteriana. Sin embargo, como ya mencio-

namos, el concepto biológico de especie *per se* tiene varios inconvenientes para ser usado como un concepto de especie en general. Dykhuizen y Green (1991) hacen el esfuerzo de acoplar este concepto de especie a bacterias con una estructura genética no totalmente clonal, como *E. coli*, pero además de que pueden existir organismos del todo clonales, la rareza y la promiscuidad en el intercambio genético entre bacterias hace que el concepto biológico de especie no pueda ser aplicado ampliamente en este grupo de organismos (Cohan, 1994b; 2002).

Cohan (2001, 2002) argumenta que el intercambio genético en bacterias no puede evitar la divergencia, dada su rareza y promiscuidad, y plantea que el aislamiento sexual es irrelevante para la evolución de divergencia permanente en el mundo bacteriano. Bajo esta visión, las especies asexuales están sujetas a su propia fuerza de cohesión, la selección natural que purga la diversidad genética dentro de las poblaciones. En este fenómeno, conocido como selección periódica, una nueva especie se forma cuando una línea asexual evoluciona hasta ocupar un nuevo nicho ecológico, de modo que la nueva especie no puede verse afectada por la selección que opera sobre los alelos adaptativos de otra población. Una vez que las poblaciones son lo suficientemente divergentes como para escapar de los eventos de selección periódica de cada una, pueden divergir permanentemente y alcanzan el estatus de especie, es decir, son ecológicamente distintas y están irreversiblemente separadas. Este autor apunta que estas propiedades universales de las especies no están presentes en las especies nombradas por los sistemáticos, sino en los ecotipos, que son organismos que ocupan el mismo nicho ecológico, cuya divergencia es purgada recurrentemente por selección natural. Las especies nombradas contienen varios ecotipos, por lo que propone que se les reconozca utilizando una clasificación latina trinomial, nombrando su género, especie y ecotipo.

La propuesta de Cohan (2002) identifica las especies como grupos monofiléticos que comparten una historia evolutiva de selección ecológica y aclara algunos de los posibles mecanismos de especiación. Sin embargo, su propuesta tiene como debilidades que la teoría evolutiva detrás de este concepto puede ponerse en duda, ya que está basada en el modelo clonal, y que no toma en cuenta que muchos caracteres adaptativos están codificados en elementos móviles y extracromosomales. La propuesta de introducir en la sistemática bacteriana el concepto de clona, complejo clonal o ecotipo dentro de una especie ya existía; se generó en el sentido de incluir ecotipos que en la actualidad se consideran especies o incluso géneros distintos, dentro de una misma especie, como el caso de muchos patógenos humanos (Lan y Reeves,

2001). Por ejemplo, *Bacillus anthracis*, *B. cereus* y *B. thuringiensis* son clonas de una misma especie (Helgason *et al.*, 2000) y la mayoría de las especies de *Shigella* son clonas dentro de *E. coli* (Pupo *et al.*, 2000).

Sin embargo, creemos que las especies, como conjunto de ecotipos, son la unidad evolutiva principal. Las clonas forman parte de la totalidad de la especie, a escala ecológica estructuran las poblaciones y a escala evolutiva contienen la diversidad genética de la especie. Las clonas o ecotipos nacen y se extinguen, así como los individuos y las especies. Creemos que la introducción de los ecotipos en la nomenclatura de las especies bacterianas es plausible, y de hecho se hace, como en el caso de serotipos, biovariedades, subespecies, etc. Sin embargo, no es válido transferirles los atributos de la categoría y el concepto de especie. Por ejemplo, dentro de la especie *R. etli* se reconocen tres ecotipos: los no simbióticos aislados del suelo, la biovariedad *phaseoli*, aislada de nódulos de frijol, y la biovariedad *mimosae*, aislada de nódulos de mimosas (Segovia *et al.*, 1991; Wang *et al.*, 1999). Segovia *et al.* (1991) demostraron que la transferencia del plásmido simbiótico de la biovar *phaseoli* convertía en simbióticos a los *R. etli* no simbióticos. Este ejemplo permite mostrar la manera en que diferentes ecotipos se relacionan entre sí y mantiene la dinámica evolutiva dentro de una especie.

Tres estudios de caso con rizobios

Diversidad de los rizobios asociados a frijoles cultivados bajo el sistema de milpa en México

Las evidencias arqueológicas, morfológicas y moleculares sugieren que el frijol cultivado evolucionó a partir de su pariente silvestre (*Phaseolus vulgaris*) y que fue domesticado en múltiples sitios de Mesoamérica y la región andina de Sudamérica hace unos 4,000 años (Singh *et al.*, 1991; Kaplan y Lynch, 1999). La llegada de los europeos a América disparó la diseminación de este cultivo hacia Europa y otras partes del mundo, y probablemente los rizobios americanos adheridos a las semillas fueron también transportados (Pérez-Ramírez *et al.*, 1998). *Rhizobium etli* bv. *phaseoli* es la especie que con mayor frecuencia se encuentra nodulando los frijoles en América y se ha postulado que las biovariedades *phaseoli* de diversas especies (*R. leguminosarum*, *R. gallicum* y *R. gardinii*) que nodulan al frijol surgieron por la transferencia lateral del plásmido simbiótico proveniente de *R. etli* (Segovia *et al.*, 1993; Amarger *et al.*, 1997; Martínez-Romero, 2003). Actualmente el frijol se cultiva en todos los continentes y tiene gran importancia alimenticia y econó-

mica (Graham y Vance, 2003) y es probable que fuera co-domesticado con el maíz en Mesoamérica, ya que tradicionalmente se siembran juntos en el agroecosistema conocido como milpa (Souza *et al.*, 1997; Martínez-Romero, 2003). A continuación presentamos los principales resultados del análisis de la estructura genética de los rizobios asociados a frijoles cultivados bajo el sistema de milpa en México.

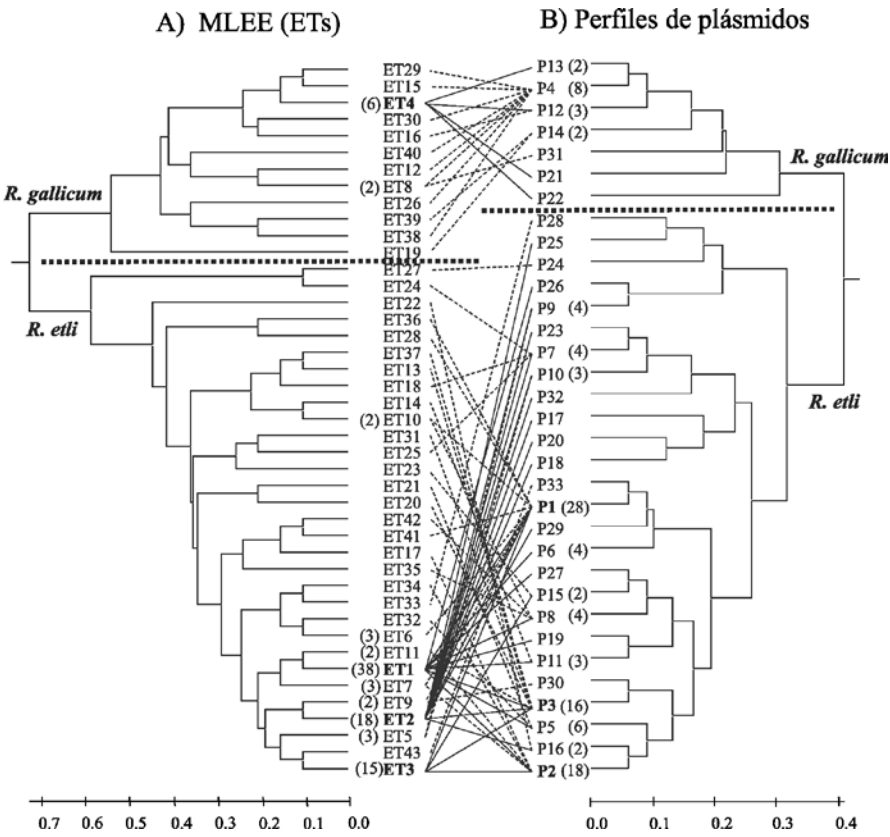
San Miguel Acuexcomac es un pueblo cercano a la ciudad de Puebla, donde se utiliza el sistema de milpa para cosechar maíz, frijol, calabaza y un gran número de hierbas que crecen en las parcelas y que se utilizan con fines alimenticios y medicinales (Souza *et al.*, 1997). Las variedades de frijol que se siembran han sido heredadas de generación en generación y son ecotipos conocidos como frijoles criollos o, en inglés, *landraces*. Con el objetivo de conocer la distribución espacial y temporal de la diversidad genética de los rizobios asociados a estas plantas de frijol, un mismo día muestreamos seis parcelas de milpa y una de las parcelas la muestreamos durante tres años consecutivos (Silva *et al.*, 1999; 2003). Para detectar la variación genética de las poblaciones de bacterias utilizamos como marcador genético la electroforesis multilocus de isoenzimas (MLEE) (véase el capítulo 18 de este libro).

Encontramos una estructura genética estable espacial y temporalmente, con una alta diversidad genética en las seis parcelas y en la parcela muestreada durante tres años (H = 0.531 y 0.501, respectivamente), y una alta dominancia ecológica (50% de los nódulos estuvieron ocupados por sólo 5 genotipos multilocus). En la muestra encontramos dos grandes grupos de genotipos separados por una distancia genética de 0.7 (figura 3). Los análisis de desequilibrio de ligamiento evidenciaron intercambio genético frecuente dentro de los dos grupos pero no entre ellos, sugiriendo la existencia de una barrera genética. Este resultado fue el primer indicio de que los grupos genéticos podrían pertenecer a dos especies de rizobios, por lo que analizamos la variación genética en una muestra de 35 aislados usando otros marcadores de cromosoma y plásmidos.

El conjunto de marcadores moleculares (incluyendo secuencia parciales de tres genes cromosomales *rrs*, *glnII* y *atpD*, y dos genes simbióticos *nifH* y *nodB*), aunados a experimentos de nodulación, mostraron que los dos grupos genéticos corresponden a dos especies previamente descritas, *R. etli* bv. *phaseoli* y *R. gallicum* bv. *gallicum* (Silva *et al.*, 2003). Además, los marcadores asociados a los plásmidos mostraron evidencia de transferencia lateral de plásmidos dentro de cada grupo pero no entre ellos (figura 3). Las conclusiones principales del análisis de la estructura genética de los rizobios asociados a

los frijoles cultivados en San Miguel fueron: 1) los frijoles de las milpas están nodulados por *R. etli* y *R. gallicum*; 2) ambas especies presentan una estructura genética poblacional epidémica, que es estable en las escalas espacial y temporal analizadas; 3) el intercambio genético es una fuerza evolutiva importante en moldear sus estructura genéticas; 4) aunque estas especies tienen

Figura 3. Dendrogramas que muestran las relaciones genéticas entre (A) genotipos cromosomales y (B) perfiles plasmídicos, así como las combinaciones cromosoma-perfil de plásmidos para 108 aislados de *R. etli* y 18 de *R. gallicum* analizados en Silva *et al.* (2003). El número de aislados está indicado en paréntesis. Las combinaciones cromosoma-perfil de plásmidos de los ETs más abundantes se indican con líneas enteras y el resto con líneas punteadas



la oportunidad ecológica para el intercambio genético, éste no es detectable, al menos para las poblaciones que ocupan los nódulos del frijol.

En este caso los análisis de genética de poblaciones nos dieron indicios de la existencia de dos especies en la muestra y los análisis filogenéticos nos permitieron establecer que se trataba de poblaciones de *R. etli* y *R. gallicum*.

Biogeografía y genética evolutiva de poblaciones de *R. gallicum* y especies relacionadas

La población de San Miguel Acuexcomac fue el primer reporte de *R. gallicum* aislado de frijol en América (Silva *et al.*, 2003). Esta especie fue descrita en Francia (Amarger *et al.*, 1997) y ha sido aislada de frijol en Austria (Sessitsch *et al.*, 1997a, 1997b), España (Herrera-Cervera *et al.*, 1999; Rodriguez-Navarro *et al.*, 2000) y Túnez (Mhamdi *et al.*, 1999, 2002). Para obtener una mejor perspectiva sobre la población de *R. gallicum* de México y ubicarla en un contexto filogenético, secuenciamos cinco genes para algunos de los aislados de San Miguel y los reportados en diversos continentes (Silva *et al.*, 2005). Obtuvimos secuencias parciales de tres genes cromosomales indispensables para el mantenimiento celular (*rrs*, *glnII* y *atpD*) y dos genes ligados al plásmido simbiótico involucrados en la interacción simbiótica con plantas (*nifH* y *nodB*).

Los análisis filogenéticos mostraron que los alelos cromosomales de *R. gallicum* forman un grupo monofilético con los de aislados de *Medicago ruthenica*, *Coronilla varia*, *Amphicarpaea trisperma* y *Gueldenstaedtia multiflora* de diferentes regiones de China (figura 4), mientras que sus alelos simbióticos están relacionados con los de algunas especies de *Sinorhizobium* asociadas a alfalfa. Con base en criterios de taxonomía clásica (hibridaciones ADN:ADN) estos aislados fueron clasificados como *R. mongolense* y *R. yanglingense* (van Berkum *et al.*, 1998; Tan *et al.*, 2001). Nuestros análisis muestran que *R. mongolense* y *R. yanglingense* pertenecen al linaje evolutivo de *R. gallicum* y que constituyen un nuevo ecotipo simbiótico que proponemos nombrar biovar *orientale* (Silva *et al.*, 2005). *R. mongolense* y *R. yanglingense* ejemplifican el error frecuente de crear nuevas especies o incluso géneros con base en caracteres ecológicos codificados por elementos genéticos móviles, un asunto ampliamente documentado para algunas bacterias patógenas de humanos (Helgason *et al.*, 2000; Pupo *et al.*, 2000; Lan y Reeves 2001).

Con base en los resultados de los análisis filogenéticos decidimos nombrar al linaje evolutivo que incluye a *R. gallicum*, *R. mongolense* y *R. yanglingense* como *R. gallicum sensu lato* (en sentido amplio), ya que *R. gallicum* fue la

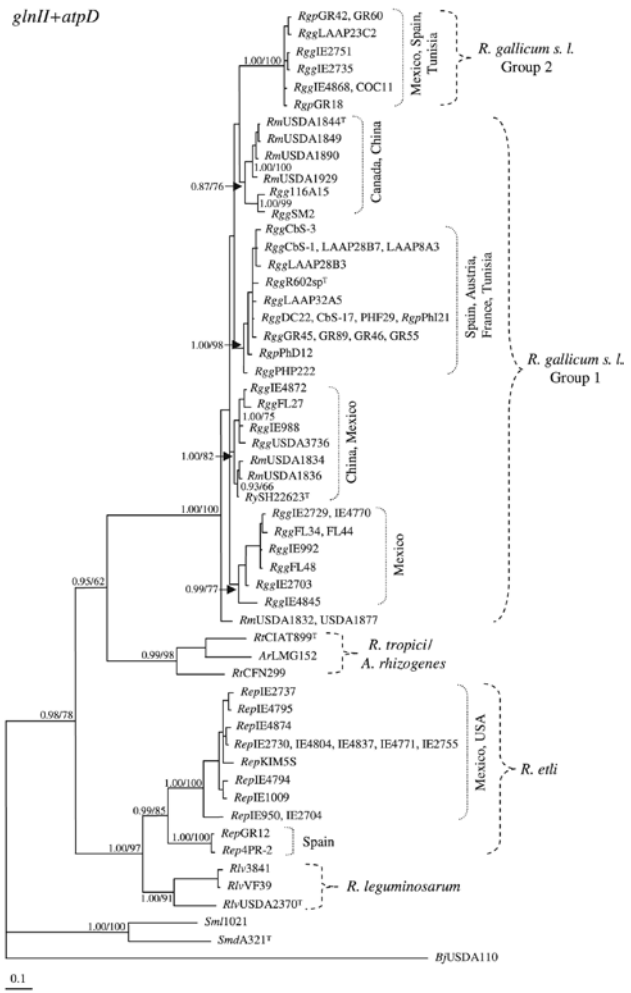
primera de las tres especies en ser descrita. En los análisis de genética de poblaciones incluimos los aislados de *R. mongolense* y *R. yanglingense* como la población de China. Tanto los análisis filogenéticos como los de genética de poblaciones basados en las secuencias de los genes *glnII* y *atpD* mostraron un patrón filogeográfico en el que las poblaciones de Francia, Austria, España y Túnez no muestran diferenciación significativa, por lo que pueden considerarse como una misma población, mientras que las de México y China mostraron diferenciación significativa con respecto a las demás.

Estos resultados indican que el flujo génico es una fuerza evolutiva que da cohesión a la especie y que sólo las poblaciones separadas a más de 10,000 km están genéticamente diferenciadas. En contraste, las secuencias del gen *rrs* no mostraron evidencia de patrón filogeográfico alguno, probablemente por los altos niveles de recombinación detectados para este gen, cuyas estimaciones están basadas en el análisis de polimorfismos de ADN. Asimismo, la inferencia filogenética mostró evidencia de transferencia lateral de alelos *rrs* entre varias especies de *Rhizobium* (Silva *et al.*, 2005). Estos resultados se unen al cúmulo de evidencias sobre la transferencia lateral del rADN entre especies bacterianas y en particular entre rizobios. Por lo tanto, sugerimos que para estudios sobre evolución, sistemática y biogeografía de bacterias es preferible utilizar como marcadores moleculares genes de mantenimiento celular que codifican para proteínas.

Los análisis filogenéticos evidenciaron la transferencia de loci simbióticos entre diversos linajes de rizobios. Sin embargo, los análisis de genética de poblaciones mostraron que la recombinación entre los diversos alelos simbióticos no es frecuente, probablemente por la presión selectiva a la que están sujetos estos genes involucrados en la interacción con los hospederos. El hecho de que cepas con fondos cromosomales monofiléticos (*R. gallicum sensu lato*) posean genotipos simbióticos divergentes (biovariedades *gallicum*, *phaseoli* y *orientale*) ejemplifica cómo una especie puede tener diferentes ecotipos sin perder su identidad (Silva *et al.*, 2005). Por esta razón, no compartimos la propuesta de Cohan (2002) sobre que los ecotipos tienen las propiedades esenciales de las especies.

En su conjunto los análisis filogenéticos y de genética de poblaciones permitieron delinear el linaje evolutivo de *R. gallicum sensu lato* y dilucidar el papel de las diferentes fuerzas evolutivas en moldear la estructura genética de este linaje, haciendo énfasis en las historias evolutivas contrastantes de los genes cromosomales de mantenimiento y los genes asociados al plásmido simbiótico, así como sus implicaciones en la sistemática bacteriana.

Figura 4. Filogenia bayesiana inferida a partir de las secuencias concatenadas de los genes *glnII* y *atpD*. En los nodos se muestran los valores de probabilidad posterior > 0.85 y los de bootstrap de máxima verosimilitud > 60%. Las cepas con haplotipos idénticos se incluyen en el mismo nodo terminal. Las cepas tipo se indican con una T. Las abreviaciones del género, especie y biovariedad de las cepas son: *Rgg* = *Rhizobium gallicum* bv. *gallicum*, *Rgp* = *R. gallicum* bv. *phaseoli*, *Rm* = *R. mongolense*, *Ry* = *R. yanglingense*, *Rt* = *R. tropici*, *Ar* = *Agrobacterium rhizogenes*, *Rep* = *R. etli* bv. *phaseoli*, *Rem* = *R. etli* bv. *mimosae*, *Rlv* = *R. leguminosarum* bv. *viciae*, *Sml* = *S. meliloti*, *Smd* = *S. medicae*, *Bj* = *Bradyrhizobium japonicum* . Figura adaptada de Silva *et al.* (2005).



Genética de poblaciones, filogenia y biogeografía de especies simpátricas de *Bradyrhizobium* que nodulan leguminosas genistoides endémicas de las Islas Canarias

Las Islas Canarias son de origen volcánico y están situadas en el océano Atlántico muy cerca de la costa noroeste de África. En este archipiélago las leguminosas endémicas de la tribu Genisteae son de gran importancia ecológica, ya que son especies dominantes clave en las comunidades vegetales de los ecosistemas canarios. En varios estudios se ha hecho patente que estas leguminosas están noduladas por bacterias del género *Bradyrhizobium* (Vinuesa *et al.*, 1998; Jarabo-Lorenzo *et al.*, 2000). Para estudiar las poblaciones de *Bradyrhizobium* aisladas de suelos de las Islas Canarias y áreas continentales cercanas (Marruecos y el sur de España), hemos seguido un esquema de análisis jerárquico utilizando varios marcadores moleculares con diferentes niveles de resolución taxonómica, que recomendamos para el análisis de poblaciones bacterianas cuyo tamaño de muestra es grande.

Después del aislamiento de las bacterias de los nódulos, lo primero que hicimos fueron *fingerprints* genómicos utilizando *primers* (véase el capítulo 17 de este libro) con afinidad por secuencias palindrómicas repetidas dispersas en el genoma de las bacterias, conocidos como Rep-PCR (de Bruijn, 1992). Esos fingerprints tienen un elevado nivel de resolución para identificar clonas -bacterias genómicamente idénticas-, pero no tiene resolución en niveles taxonómicos superiores. A partir de la identificación de aislados pertenecientes a una misma clona, se seleccionaron aislados representativos de la variación genética para el siguiente análisis de tipificación. El análisis del polimorfismo de fragmentos de restricción (RFLP; véase el capítulo 18 de este libro) de amplificaciones del gen que codifica para la 16S rARN y del espaciador intergénico (ITS) 16S-23S rADN mostró que la mayoría de los aislados forman un grupo independiente, para el cual no existían cepas o especies descritas, cercano al grupo formado por aislados caracterizados como *B. japonicum* (Vinuesa *et al.*, 1998, 1999).

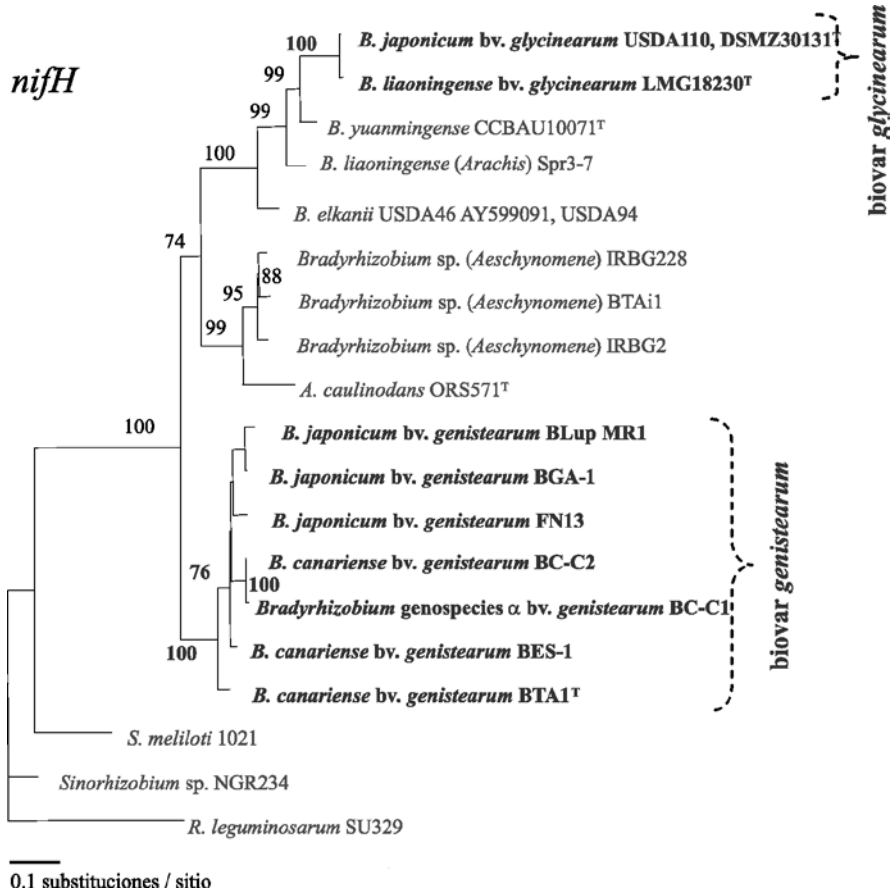
Otros aislados de Canarias mostraron RFLPs idénticos a los de *B. japonicum* y un aislado tuvo un patrón diferente a los dos anteriores. Estos resultados dieron los primeros indicios de que la mayoría de los aislados (70% de los nódulos) pertenecían a una nueva especie, que recientemente describimos como *B. canariense* (Vinuesa *et al.*, 2005a, 2005c). Experimentos de nodulación mostraron que los aislados de las leguminosas genistoides eran incapaces de

nodular soya (*Glycine* spp.) y que las cepas aisladas de soya (como *B. japonicum* y *B. elkanii*) eran incapaces de nodular a las leguminosas genistoides, lo cual dio los primeros indicios de que los aislados de Canarias representaban una nueva biovariedad o ecotipo nodulador (Vinuesa *et al.*, 2005a, 2005c).

Las electroforesis de enzimas (MLEE) fueron el siguiente marcador molecular que analizamos para conocer la dinámica poblacional de *B. canariense*. Los análisis de genética de poblaciones mostraron que no existe diferenciación genética significativa entre las poblaciones insulares y continentales de *B. canariense*, sugiriendo la existencia de flujo genético entre ellas. Los análisis de desequilibrio de ligamiento (véase el capítulo 9 de este libro) mostraron equilibrio para la población de *B. canariense* y desequilibrio al analizarla junto con los aislados de *B. japonicum*, indicando la posible existencia de una barrera al intercambio genético entre estas dos especies simpátricas (Vinuesa *et al.*, 2005c). Los marcadores moleculares que han aportado la información concluyente sobre el estatus taxonómico y evolutivo de estas bacterias han sido las secuencias nucleotídicas de cuatro loci involucrados en el mantenimiento celular (*atpD*, *glnII*, *recA* e ITS 16S-23S rADN) y dos genes involucrados en la simbiosis (*nifH* y *nodC*). Los análisis filogenéticos basados en los genes de mantenimiento muestran que entre los aislados de las leguminosas genistoides hay cuatro linajes principales: la nueva especie que describimos *B. canariense*, la especie descrita *B. japonicum*, y dos linajes referidos como las genoespecies α y β (Vinuesa y Silva 2004; Vinuesa *et al.*, 2005a, 2005c). Estos linajes forman grupos monofiléticos son un alto soporte en las filogenias de los genes individuales como en la filogenia de especies generada a partir de las secuencias congruentes concatenadas *glnII+recA*.

Por otro lado, las filogenias derivadas de los dos genes involucrados en la simbiosis forman un grupo monofilético para los aislados de los cuatro linajes cromosomales (figura 5). Este resultado evidencia eventos de transferencia lateral de loci simbióticos entre fondos cromosomales pertenecientes a linajes divergentes. A esta nueva biovariedad noduladora de leguminosas genistoides la hemos nombrado *genistearum* (Vinuesa *et al.*, 2005a, 2005c). Los análisis de genética de poblaciones basados en las secuencias de los loci de mantenimiento refuerzan las conclusiones derivadas del análisis de isoenzimas, mostrando nula diferenciación entre las poblaciones de *B. canariense* insulares y continentales, una muy significativa diferenciación entre las poblaciones simpátricas de *B. canariense* y *B. japonicum*, y recombinación significativa entre los miembros de cada población, pero no entre especies. Este último resultado fue también evidenciado con un análisis de *split decomposition* (Huson, 1998;

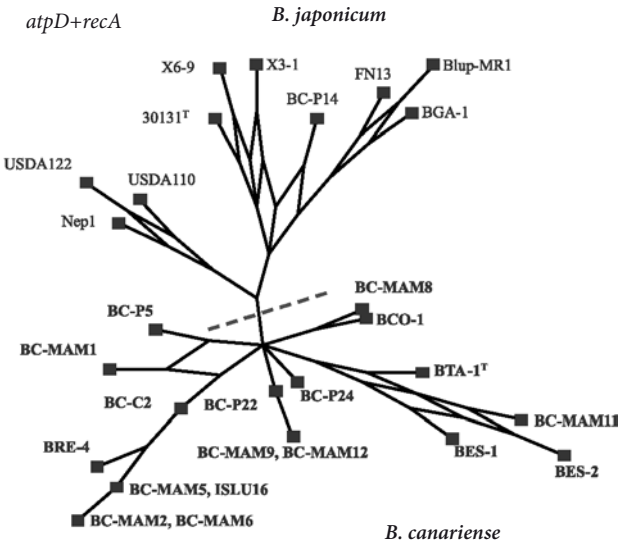
Figura 5. Filogenia de máxima verosimilitud para secuencias *nifH* de *Bradyrhizobium*. En los nodos se muestran los valores de bootstrap > 70%. Con base en la correlación entre clados *nifH* y especificidad por hospederos se proponen por primera vez dos biovariedaes para el género *Bradyrhizobium*: la biovariedad *glycinearum* de aislados asociados a soya y la biovariedad *genistearum* asociada a leguminosas genistoides. Figura publicada en el estudio de Vinuesa *et al.* (2005a).



véase también el artículo 4 de este libro)) que mostró evolución reticulada dentro de las poblaciones pero no entre especies (figura 6).

El conjunto de resultados muestran que la migración y la recombinación son importantes fuerzas de cohesión genética que mantienen la identidad de

Figura 6. Análisis de *split decomposition* para las secuencias concatenadas *atpD* y *recA*. En la gráfica se observan los patrones de evolución reticulada dentro de las poblaciones de *B. canariense* y *B. japonicum*. Sin embargo, no se observa ningún indicio de reticulación que involucre a las dos especies. Nótese que la gráfica está representada con vértices iguales (no proporcionales a la distancia genética). La línea punteada indica la rama más larga cuando la gráfica se hace proporcional a la distancia genética. Tomado de Vinuesa *et al.* (2005c).



las especies; existe una barrera para el intercambio genético de genes de mantenimiento entre especies simpátricas, y que, sin embargo, la transferencia lateral de genes simbióticos entre especies ha ocurrido. La distribución geográfica de *B. canariense* bv. *genistearum* y *B. japonicum* bv. *genistearum* incluye a las Islas Canarias y áreas continentales cercanas, Norteamérica y probablemente se extiende a otras regiones (Jarabo-Lorenzo *et al.*, 2003; Vinuesa *et al.*, 2005c). Creemos que estas bacterias pueden estar siendo transportadas junto con las partículas de polvo que surgen de África y son llevadas por las corrientes atmosféricas a diferentes partes del mundo (Griffin *et al.*, 2002; Prospero, 2004).

Conclusiones

Todo parece indicar que los genomas bacterianos evolucionan por módulos: genes organizados en operones, cromosomas con regiones variables móviles

(casetes o islas) y arquitecturas genómicas compuestas por varios replicones (cromosomas y plásmidos). Esto produce genomas mosaico, con unidades intercambiables que pueden estar o no presentes en los diferentes individuos de una misma especie (Lan y Reeves, 2000). Además del recambio de información genética entre individuos de una especie, la transferencia lateral de información genética entre individuos de especies, géneros o incluso dominios diferentes está ampliamente documentada, en particular para genes que confieren fuertes ventajas selectivas.

La mayoría de los estudios sobre evolución bacteriana ha analizado poblaciones de bacterias patógenas de humanos. En nuestra opinión el estudio de los rizobios ha ampliado la perspectiva acerca de la evolución bacteriana. Dado que la mayor parte del ciclo de vida de los rizobios ocurre en el suelo y que sólo una fracción de los individuos de las poblaciones entra en simbiosis, podemos estudiar diversos aspectos del efecto selectivo del hospedero sobre el genoma accesorio de los rizobios y la evolución del genoma central de una bacteria de vida libre. La ganancia o la pérdida de información simbiótica produce ecotipos con la capacidad de ocupar nichos específicos (los nódulos de una leguminosa particular) y dentro de una especie pueden existir diferentes ecotipos a la vez. La información simbiótica puede ser transferida entre individuos de una especie o incluso géneros o familias diferentes.

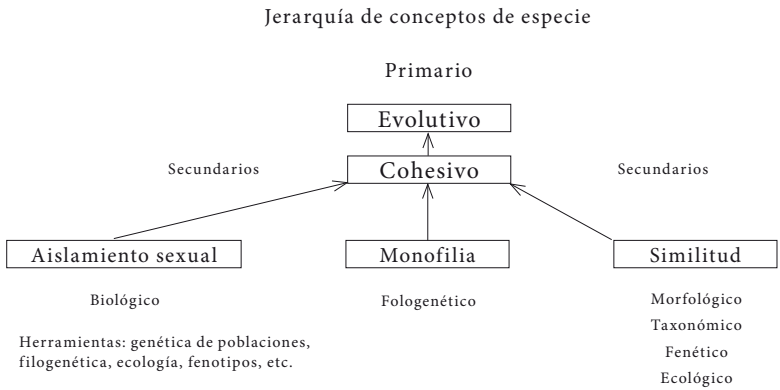
Del estudio de los rizobios nos queda una visión de las especies bacterianas como entidades cohesivas donde la recombinación homóloga dentro de las poblaciones, la migración y el flujo génico juegan papeles importantes, incluso en la escala global. El examen de las comunidades microbianas asociadas a las partículas de polvo transportadas entre continentes por tormentas originadas en Asia y África parecen estrategias prometedoras para descubrir las vías de dispersión de los rizobios, como ya se ha documentado para otras bacterias y hongos del suelo (Griffin *et al.*, 2002; Prospero, 2004).

En nuestra experiencia la delimitación de especies bacterianas debe partir de un buen muestreo de poblaciones, así como de un esquema jerárquico de análisis de la variación genética donde se utilicen diferentes marcadores moleculares y fenotípicos que permitan conocer la variación de loci de mantenimiento y adaptativos, combinando análisis filogenéticos y de genética de poblaciones para resolver problemas interrelacionados sobre ecología evolutiva y sistemática bacteriana. Aunque el gen *rrs* (16S rARN) es el marcador genético más utilizado, cada vez queda más claro que por su frecuencia de recombinación y número variable de copias por genoma, no es el locus más adecuado para estudios filogenéticos y ecológicos. Es preferible analizar genes

de mantenimiento que codifican para proteínas, ya que contienen mayor información y existen más herramientas para analizar sus procesos de evolución molecular (véase el capítulo 1 en este libro).

Los métodos clásicos utilizados en la delimitación de especies bacterianas son útiles para evaluar la similitud entre cepas, pero no es válido imponer puntos de corte arbitrarios para delimitar especies. Creemos que el concepto de especie evolutivo es el que mejor puede manejar la enorme diversidad del mundo vivo (figura 7). Este concepto tiene la capacidad de incluir el conocimiento que tenemos sobre la ecología y evolución bacteriana, en particular, el papel fundamental del intercambio genético dentro y entre especies. Tenemos que aceptar que las especies no son estáticas ni sistemas cerrados. Es importante incorporar el pensamiento poblacional a la sistemática bacteriana para poder estudiar en un marco más amplio la variación genética, las relaciones filogenéticas y los caracteres y mecanismos ecológicos que dan cohesión a las especies.

Figura 7. Se plantea que el concepto evolutivo de especie sea el concepto primario y que los demás conceptos de especie (basados en los criterios de similitud, monofilia y aislamiento sexual) se utilicen como conceptos secundarios para aportar información que sustente al concepto evolutivo. Se propone que las mejores herramientas para buscar las fuerzas de cohesión y delinear a las especies son la genética de poblaciones y la filogenética, así como estudios ecológicos y fenotípicos. Adaptado de Mayden (1997)



Bibliografía

Amann, R., y Kuhl, M.1998. *In situ* methods for assessment of microorganisms y their activities. *Curr. Opin. Microbiol.* 1:352-358.

Amann, R. I., Ludwig, W., y Schleifer, K. H. 1995. Phylogenetic identification y *in situ* detection of individual microbial cells without cultivation. *Microbiol. Rev.* 59:143-169.

Amarger, N., Macheret, V., y Laguerre, G. 1997. *Rhizobium gallicum* sp. nov. y *Rhizobium giardinii* sp. nov., from *Phaseolus vulgaris* nodules. *Int. J. Syst. Bacteriol.* 47:996-1006.

Asai, T., Zaporojets, D., Squires, C., y Squires, C. L. 1999. An *Escherichia coli* strain with all chromosomal rARN operons inactivated:complete exchange of rARN genes between bacteria. *Proc. Natl. Acad. Sci.* USA 96:1971-1976.

Brewin, N. J. 1998. Tisue y cell invasion by *Rhizobium*:the structure y development of infection threads y symbiosomes. Pages 417-429 in:The *Rhizobiaceae*. H. P. Spaink, A. Kondorosi, y P. J. J. Hooykaas (eds.). Kluwer Academic Publishers, Dordrecht.

Brochier, C., Philippe, H., y Moreira, D. 2000. The evolutionary history of ribosomal protein RpS14:horizontal gene transfer at the heart of the ribosome. *Trends Genet.* 16:529-533.

Brockwell, J., Bottomley, P. J., y Thies, J. E. 1995. Manipulation of rhizobia microflora for improving legume productivity y soil fertility:a critical assessment. *Plant Soil* 174:143-180.

Brom, S., Girard, L., García-de los Santos, A., Sanjuan-Pinilla, J. M., Olivares, J., y Sanjuan, J. 2002. Conservation of plasmid-encoded traits among bean-nodulating *Rhizobium* species. *Appl. Environ. Microbiol.* 68:2555-2561.

Brown, E. W., Kotewicz, M. L., y Cebula, T. A. 2002. Detection of recombination among *Salmonella enterica* strains using the incongruence length difference test. *Mol. Phylogenet. Evol.* 24:102-120.

Chen, W.-M., Laevens, S., Lee, T.-M., Coenye, T., de Vos, P., Mergeay, M., y Vandamme, P. 2001. *Ralstonia taiwanensis* sp. nov., isolated from root nodules of *Mimosa* species y sputum of cystic fibrosis patient. *Int. J. Syst. Evol. Microbiol.* 51:1729-1735.

Chen, W. M., Moulin, L., Bontemps, C., Vandamme, P., Bena, G., y Boivin-Masson, C. 2003. Legume symbiotic nitrogen fixation by beta-proteobacteria is widespread in nature. *J. Bacteriol.* 185:7266-7272.

Cheng, T. C. 1991. Is parasitism symbiosis? A definition of terms y the evolution of concepts. Pages 15-36 in:Parasite host associations. Coexistence or conflict? C. A. Tof, A. Aeschlimann, y L. Bolis (eds.). Oxford University Press, Oxford.

Claridge, M. F., Dawah, H. A., y Wilson, M. R. 1997. Species:the units of biodiversity, Chapman y Hall, London.

Cohan, F. M. 1994a. The effects of rare but promiscuous genetic exchange on evolutionary divergence in prokaryotes. *Am. Nat.* 143:965-986.

Cohan, F. M. 1994b. Genetic exchange y evolutionary divergence in prokaryotes. *Trends Ecol. Evol.* 9:175-180.

Cohan, F. M. 1996. The role of genetic exchange in bacterial evolution. *ASM News* 62:631-636.

Cohan, F. M. 2002. What are bacterial species? *Annu. Rev. Microbiol.* 56:457-487.

Daubin, V., Moran, N. A., y Ochman, H. 2003. Phylogenetics y the cohesion of bacterial genomes. *Science* 301.

de Bruijn, F. J. 1992. Use of repetitive (repetitive extragenic palindromic y enterobacterial intergeneric consensus) sequences y the polymerase chain reaction to fingerprint the genomes of *Rhizobium meliloti* isolates y other soil bacteria. *Appl. Environ. Microbiol.* 58:2180-2187.

Doolittle, R. F. 2002. Microbial genomes multiply. *Nature* 416:697-700.

Doolittle, R. F., Feng, D. F., Anderson, K. L., y Alberro, M. R. 1990. A naturally occurring horizontal gene transfer from a eukaryote to a prokaryote. *J. Mol. Evol.* 31:383-388.

Dowson, C. G., Hutchison, A., Brannigan, J. A., George, R. C., Hansman, D., Linares, J., Tomasz, A., Smith, J. M., y Spratt, B. G. 1989. Horizontal transfer of penicillin-binding protein genes in penicillin-resistant clinical isolates of *Streptococcus pneumoniae*. *Proc. Natl. Acad. Sci. USA* 86:8842-6.

Dykhuizen, D. E., y Green, L. 1991. Recombination in *Escherichia coli* y the definition of biological species. *J. Bacteriol.* 173:7257-7268.

Dykhuizen, D. E., Polin, D. S., Dunn, J. J., Wilske, B., Preac-Musric, V., y Dattwyler, R. J. 1992. *Borrelia borgdorferi* is clonal:implications for taxonomy y vaccine development. *Proc. Natl. Acad. Sci. USA* 90:10162-10167.

Eardly, B. D., Wang, F.-S., y van Berkum, P. 1996. Corresponding 16S rARN gene segments in Rhizobiaceae y *Aeromonas* yield discordant phylogenies. *Plant Soil* 186:69-74.

Feil, E. J., y Spratt, B. G. 2001. Recombination y the population structures of bacterial pathogens. *Annu. Rev. Microbiol.* 55:561-590.

Freiberg, C., Fellay, R., Bairoch, A., Broughton, W. J., Rosenthal, A., y Perret, X. 1997. Molecular basis of symbiosis between *Rhizobium* y legumes. *Nature* 387:394-401.

Freiberg, C., Perret, X., Broughton, W. J., y Rosenthal, A. 1996. Sequencing the 500-kb GC-rich symbiotic replicon of *Rhizobium* sp. NGR234 using dye terminators y a thermostable "sequenase":a beginning. *Genome. Res.* 6:590-600.

Galibert, F., Finan, T. M., Long, S. R., Puhler, A., Abola, P., Ampe, F., Barloy-Hubler, F., Barnett, M. J., Becker, A., Boistard, P., Bothe, G., Boutry, M., Bowser, L., Buhrmester, J., Cadieu, E., Capela, D., Chain, P., Cowie, A., Davis, R. W., Dreano, S., Federspiel, N. A., Fisher, R. F., Gloux, S., Godrie, T., Goffeau, A., Golding, B., Gouzy, J., Gurjal, M., HeARNndez-Lucas, I., Hong, A., Huizar, L., Hyman, R. W., Jones, T., Kahn, D., Kahn, M. L., Kalman, S., Keating, D. H., Kiss, E., Komp, C., Lelaure, V., Masuy, D., Palm, C., Peck, M. C., Pohl, T. M., Portetelle, D., Purnelle, B., Ramsperger, U., Surzycki, R., Thebault, P., Vandenbol, M., Vorholter, F. J., Weidner, S., Wells, D. H., Wong, K., Yeh, K. C., y Batut, J. 2001. The composite genome of the legume symbiont *Sinorhizobium meliloti*. *Science* 293:668-672.

Gogarten, J. P., Doolittle, W. F., y Lawrence, J. G. 2002. Prokaryotic evolution in light of gene transfer. *Mol. Biol. Evol.* 19:2226-2238.

Gogarten, J. P., Murphey, R. D., y Olendzenski, L. 1999. Horizontal gene transfer: pitfalls y promises. *Biol. Bull.* 196:359-362.

González, V., Bustos, P., Ramírez-Romero, M. A., Medrano-Soto, A., Salgado, H., Hernández-Gonzáles, I., Hernández-Celis, J. C., Quintero, V., Moreno-Hagelsieb, G., Girard, L., Rodríguez, O., Flores, M., Cevallos, M. A., Collado-Vides, J., Romero, D., y Dávila, G. 2003. The mosaic structure of the symbiotic plasmid of *Rhizobium etli* CFN42 y its relation to other symbiotic genome compartments. *Genome. Biol.* 4:R36.

Gordon, D. M., Wexler, M., Reardon, T. B., y Murphy, P. J. 1995. The genetic structure of *Rhizobium* populations. *Soil Biology y Biochemistry* 27:491-499.

Graham, P. H., y Vance, C. P. 2003. Legumes:Importance y constraints to greater use. *Plant Physiol.* 131:872-877.

Griffin, D. W., Kellogg, C. A., Garrison, V. H., y Shinn, E. A. 2002. The global transport of dust - An intercontinental river of dust, microorganisms y toxic chemicals flows through the Earth's atmosphere. *Am. Sci.* 90:228-235.

Hashimoto, J. G., Stevenson, B. S., y Schmidt, T. M. 2003. Rates y consequences of recombination between rARN operons. *J. Bacteriol.* 185:966-972.

Helgason, E., Okstad, O. A., Caugant, D. A., Johansen, H. A., Fouet, A., Mock, M., Hegna, I., y Kolsto. 2000. *Bacillus anthracis, Bacillus cereus,* y *Bacillus thuringiensis*--one species on the basis of genetic evidence. *Appl. Environ. Microbiol.* 66:2627-2630.

Herrera-Cervera, J. A., Caballero-Mellado, J., Laguerre, G., Tichy, H. V., Requena, N., Amarger, N., Martínez-Romero, E., Olivares, J., y Sanjuan, J. 1999. At least five rhizobial species nodulate *Phaseolus vulgaris* in a Spanish soil. *FEMS Microbiol. Ecol.* 30:87-97.

Holmes, E. C., Urwin, R., y Maiden, M. C. 1999. The influence of recombination on the population structure y evolution of the human pathogen *Neisseria meningitidis*. *Mol. Biol. Evol.* 16:741-749.

Hooykaas, P. J., Snijdewint, F. G., y Schilperoort, R. A. 1982. Identification of the Sym plasmid of *Rhizobium leguminosarum* strain 1001 y its transfer to y expression in other rhizobia y *Agrobacterium tumefaciens*. *Plasmid* 8:73-82.

Hooykaas, P. J. J., Klapwijk, P. M., Nuti, M. P., Schilperoort, R. A., y Rorsch, A. 1977. Transfer of the *Agrobacterium tumefaciens* Ti-plasmid to avirulent agrobacteria y to *Rhizobium* ex planta. *J. Gen. Microbiol.* 98:477-484.

Hooykaas, P. J. J., y Schilperoort, R. A. 1984. The molecular genetics of crown gall tumorigenisis. *Adv. Genet.* 22:209-283.

Hull, D. L. 1997. The ideal species concept -y why we can't get it. Pages in:Species: the units of biodiversity. M. F. Claridge, H. A. Dawah, y M. R. Wilson (eds.). Chapman y Hall, London.

Huson, D. H. 1998. SplitsTree:analyzing y visualizing evolutionary data. *Bioinformatics* 14:68-73.

Jain, R., Rivera, M. C., y Lake, J. A. 1999. Horizontal gene transfer among genomes: the complexity hypothesis. *Proc. Natl. Acad. Sci. USA* 96:3801-3806.

Jarabo-Lorenzo, A., Pérez-Galdona, R., Donate-Correa, J., Rivas, R., Velázquez, E., Hernández, M., Temprano, F., Martínez-Molina, E., Ruiz-Argüeso, T., y León-Barrios, M. 2003. Genetic diversity of bradyrhizobial populations from diverse geographic origins that nodulate *Lupinus* spp. y *Ornithopus* spp. *Syst. Appl. Microbiol.* 26:611-623.

Jarabo-Lorenzo, A., Velazquez, E., Perez-Galdona, R., Vega-HeARNndez, M. C., Martinez-Molina, E., Mateos, P. E., Vinuesa, P., Martinez-Romero, E., y Leon-Barrios, M. 2000. Restriction fragment length polymorphism analysis of 16S rADN y low molecular weight ARN profiling of rhizobial isolates from shrubby legumes endemic to the Canary islands. *Syst. Appl. Microbiol.* 23:418-425.

Kaneko, T., Nakamura, Y., Sato, S., Asamizu, E., Kato, T., Sasamoto, S., Watanabe, A., Idesawa, K., Ishikawa, A., Kawashima, K., Kimura, T., Kishida, Y., Kiyokawa, C., Kohara, M., Matsumoto, M., Matsuno, A., Mochizuki, Y., Nakayama, S., Nakazaki, N., Shimpo, S., Sugimoto, M., Takeuchi, C., Yamada, M., y Tabata, S. 2000. Complete genome structure of the nitrogen-fixing symbiotic bacterium *Mesorhizobium loti*. *ADN Res.* 7:331-338.

Kaneko, T., Nakamura, Y., Sato, S., Minamisawa, K., Uchiumi, T., Sasamoto, S., Watanabe, A., Idesawa, K., Iriguchi, M., Kawashima, K., Kohara, M., Matsumoto, M., Shimpo, S., Tsuruoka, H., Wada, T., Yamada, M., y Tabata, S. 2002. Complete genomic sequence of nitrogen-fixing symbiotic bacterium *Bradyrhizobium japonicum* USDA110. *ADN Res.* 9:189-197.

Kaplan, L., y Lynch, T. F. 1999. *Phaseolus* (Fabaceae) in archeology:AMS radiocarbon dates y their significance for pre-Columbian agriculture. *Econ. Bot.* 53:261-272.

Laguerre, G., Geniaux, E., Mazurier, S. I., Rodríguez Casartelli, R., y Amarger, N. 1992. Conformity y diversity among field isolates of *Rhizobium leguminusarum* bv. *viciae*, bv. *trifolii*, y bv. *phaseoli* revealed by ADN hybridization using chromosome y plasmid probes. *Can. J. Microbiol.* 39:412-419.

Laguerre, G., Nour, S. M., Macheret, V., Sanjuan, J., Drouin, P., y Amarger, N. 2001. Classification of rhizobia based on *nodC* y *nifH* gene analysis reveals a close phylogenetic relationship among *Phaseolus vulgaris* symbionts. *Microbiology* 147:981-993.

Lan, R., y Reeves, P. 2001. When does a clone deserve a name? A perspective on bacterial species based on population genetics. *Trends Microbiol.* 9:419-424.

Lan, R., y Reeves, P. R. 1998. Recombination between rARN operons created most of the ribotype variation observed in the seventh pandemic clone of *Vibrio cholerae*. *Microbiology* 144:1213-1221.

Lan, R. T., y Reeves, P. R. 2000. Intraspecies variation in bacterial genomes:the need for a species genome concept. *Trends Microbiol.* 8:396-401.

Lawrence, J. G. 1999. Gene transfer, speciation, y the evolution of bacterial genomes. *Curr. Opin. Microbiol.* 2:519-523.

Lawrence, J. G. 2000. Clustering of antibiotic resistance genes:beyond the selfish operon. *ASM News* 66:281-286.

Lawrence, J. G. 2002. Gene transfer in bacteria:speciation without species? *Theor. Popul. Biol.* 61:449-460.

Lawrence, J. G., y Hendrickson, H. 2003. Lateral gene transfer:when will adolescence end? *Mol. Microbiol.* 50:739-749.

Lawrence, J. G., y Roth, J. R. 1996. Selfish operons:horizontal transfer may drive the evolution of gene clusters. *Genetics* 143:1843-1860.

Levin, B. R., y Bergstrom, C. T. 2000. Bacteria are different:observations, interpretations, speculations, y opinions about the mechanisms of adaptive evolution in prokaryotes. *Proc. Natl. Acad. Sci. USA* 97:6981-6985.

Li, W.-H. 1997. *Molecuar evolution*, Sinauer Associates, Inc., Publishers, Massachusetts.

Long, S. R. 1989. *Rhizobium*-legume nodulation:Life together in the underground. *Cell* 56:203-214.

Louvrier, P., Laguerre, G., y Amarger, N. 1996. Distribution of symbiotic genotypes in *Rhizobium leguminosarum* biovar *viciae* populations isolated directly from soils. *Appl. Environ. Microbiol.* 62:4202-4205.

Ludwig, W., y Schleifer, K. H. 1994. Bacterial phylogeny based on 16S y 23S rARN sequence analysis. *FEMS Microbiol. Rev.* 15:155-173.

Madigan, M. T., Martiniko, J. M., y Parker, J. 2000. Brock biology of microorganisms, Prentice-Hall, Inc., New Jersey.

Maggi Solcà, N., BeARNsconi, M. V., Valsangiacomo, C., Van Doorn, L. J., y Piffaretti, J. C. 2001. Population genetics of *Helicobacter pylori* in the southern part of Switzerly analysed by sequencing of four housekeeping genes (*atpD*, *glnA*, *scoB* y *recA*), y by *vacA*, *cagA*, *iceA* y *IS605* genotyping. *Microbiology* 147:1693-1707.

Majewski, J. 2001. Sexual isolation in bacteria. *FEMS Microbiol. Lett.* 199:161-169.

Martínez, E. 1994. Recent developments in *Rhizobium* taxonomy. *Plant Soil* 161:11-20.

Martínez, E., Palacios, R., y Sanchez, F. 1987. Nitrogen-fixing nodules induced by *Agrobacterium tumefaciens* harboring *Rhizobium phaseoli* plasmids. *J. Bacteriol.* 169:2828-2834.

Martínez-Murcia, A. J., Benlloch, S., y Collins, M. D. 1992. Phylogenetic interrelationships of members of the genera *Aeromonas* y *Plesiomonas* as determined by 16S ribosomal ADN sequencing:lack of congruence with results of ADN-ADN hybridizations. *Int. J. Syst. Bacteriol.* 42:412-421.

Martínez-Romero, E. 2003. Diversity of *Rhizobium-Phaseolus vulgaris* symbiosis: overview y perspectives. *Plant Soil* 252:11-23.

Matic, I., Taddei, F., y Radman, M. 1996. Genetic barriers among bacteria. *Trends Microbiol.* 4:69-72.

Mayden, R. L. 1997. A hierarchy of species concepts:the denouement in the saga of the species problem. Pages in:Species:the units of biodiversity. M. F. Claridge, H. A. Dawah, y M. R. Wilson (eds.). Chapman y Hall, London.

Maynard Smith, J. 1995. Do bacteria have population genetics? Pages 1-12 in:Population genetics of bacteria. E. A. Boumberg (ed.) Cambridge University Press.

Maynard Smith, J., Smith, N. H., O'Rourke, M., y Spratt, B. G. 1993. How clonal are bacteria? *Proc. Natl. Acad. Sci. USA* 90:4384-4388.

Mayr, E. 1970. Populations, species, y evolution, Belknap Press of Harvard University, Massachusetts.

Mhamdi, R., Jebara, M., Aouani, M. E., Ghrir, R., y Mars, M. 1999. Genotypic diversity y symbiotic effectiveness of rhizobia isolated from root nodules of *Phaseolus vulgaris* L. grown in Tunisian soils. *Biol. Fertil. Soils* 28:313-320.

Mhamdi, R., Laguerre, G., Aouani, M. E., Mars, M., y Amarger, N. 2002. Different species y symbiotic genotypes of field rhizobia can nodulate *Phaseolus vulgaris* in Tunisian soils. *FEMS Microbiol. Ecol.* 41:77-84.

Moulin, L., Munive, A., Dreyfus, B., y Boivin-Masson, C. 2001. Nodulation of legumes by members of the beta-subclass of Proteobacteria. *Nature* 411:948-950.

Mylvaganam, S., y Dennis, P. P. 1992. Sequence heterogeneity between the two genes encoding 16S rARN from the halophilic archaebacterium *Haloacula marismortui*. *Genetics* 130:399-410.

Nei, M., y Kumar, S. 2000. Molecular evolution y phylogenetics, Oxford University Press, New York.

Nelson, K. E., Clayton, R. A., Gill, S. R., Gwinn, M. L., Dodson, R. J., Haft, D. H., Hickey, E. K., Peterson, J. D., Nelson, W. C., Ketchum, K. A., McDonald, L., Utterback, T. R., Malek, J. A., Linher, K. D., Garrett, M. M., Stewart, A. M., Cotton, M. D., Pratt, M. S., Phillips, C. A., Richardson, D., Heidelberg, J., Sutton, G. G., Fleischmann, R. D., Eisen, J. A., Fraser, C. M., y et al. 1999. Evidence for lateral gene transfer between Archaea y bacteria from genome sequence of *Thermotoga maritima*. *Nature* 399:323-329.

Parker, M. A. 2001. Case of localized recombination in 23S rARN genes from divergent bradyrhizobium lineages associated with neotropical legumes. *Appl. Environ. Microbiol.* 67:2076-2082.

Pérez-Ramírez, N. O., Rogel, M. A., Wang, E., Castellanos, J. Z., y Martínez-Romero, E. 1998. Seeds of *Phaseolus vulgaris* bean carry *Rhizobium etli*. *FEMS Microbiol. Ecol.* 26:289-296.

Perna, N. T., Plunkett, G., 3rd, Burland, V., Mau, B., Glasner, J. D., Rose, D. J., Mayhew, G. F., Evans, P. S., Gregor, J., Kirkpatrick, H. A., Posfai, G., Hackett, J., Klink, S., Boutin, A., Shao, Y., Miller, L., Grotbeck, E. J., Davis, N. W., Lim, A., Dimalanta, E. T., Potamousis, K. D., Apodaca, J., Anantharaman, T. S., Lin, J., Yen, G., Schwartz, D. C., Welch, R. A., y Blattner, F. R. 2001. Genome sequence of enterohaemorrhagic *Escherichia coli* O157:H7. *Nature* 409:529-533.

Prospero, J. M. 2004. Intercontinental transport of viable fungi y bacteria from Africa to the Caribbean with soil dust. Pages in:Biological Resources y Migration. D. Werner (ed.). Springer Verlag, Berlin.

Pupo, G. M., Lan, R., y Reeves, P. R. 2000. Multiple independent origins of *Shigella* clones of *Escherichia coli* y convergent evolution of many of their characteristics. *Proc. Natl. Acad. Sci. USA* 97:10567-10572.

Rodriguez-Navarro, D. N., Buendia, A. M., Camacho, M., Lucas, M. M., y Santamaria, C. 2000. Characterization of Rhizobium spp. bean isolates from South-West Spain. *Soil Biol. Biochem.* 32:1601-1613.

Rogel, M. A., Hernández-Lucas, I., Kuykendall, L. D., Balkwill, D. L., y Martínez Romero, E. 2001. Nitrogen-fixing nodules with *Ensifer adherence* harboring *Rhizobium tropici* symbiotic plasmids. *Appl. Environ. Microbiol.* 67:3264-3268.

Romero, D., y Brom, S. 2004. The symbiotic plasmids of the *Rhizobiaceae*. En: Plasmid Biology. B. E. Funnell y G. J. Philips (eds.). ASM press, Washington. Páginas 271-290.

Sacchi, C. T., Whitney, A. M., Reeves, M. W., Mayer, L. W., y Popovic, T. 2002. Sequence diversity of *Neisseria meningitidis* 16S rARN genes y use of 16S rARN gene sequencing as a molecular subtyping tool. *J. Clin. Microbiol.* 40:4520-4527.

Sadowsky, M. J., y Graham, P. H. 1998. Soil biology of the *Rhizbiaceae*. En: The Rhizobiaceae. H. P. Spaink, A. Kondorosi, y P. J. J. Hooykaas (eds.). Kluwer Academic Publishers, Dordrecht. Páginas 155-172.

Sadowsky, M. J., y Graham, P. H. 2002. The prokaryotes:an evolving electronic resource for the microbiological community. Springer Verlag.

Sakar, S. F., y Guttman, D. S. 2004. Evolution of the core genome of *Pseudomonas syringae*, a highly clonal, endemic plant apthogen. *Appl. Environ. Microbiol.* 70:1999-2012.

Salyers, A. A., Shoemaker, N. B., Stevens, A. M., y Li, L. Y. 1995. Conjugative transposons:an unusual y diverse set of integrated gene transfer elements. *Microbiol. Rev.* 59:579-590.

Sawada, H., Kuykendall, L. D., y Young, J. M. 2003. Changing concepts in the systematics of bacterial nitrogen-fixing legume symbionts. *J. Gen. Appl. Microbiol.* 49:155-179.

Schofield, P. R., Gibson, A. H., Dudman, W. F., y Watson, J. M. 1987. Evidence for genetic exchange y recombination of *Rhizobium* symbiotic plasmid in a soil population. *Appl. Environ. Microbiol.* 53:2942-2947.

Schouls, L. M., Schot, C. S., y Jacobs, J. A. 2003. Horizontal transfer of segments of the 16S rARN genes between species of the *Streptococcus anginosus* group. *J. Bacteriol.* 185:7241-7246.

Segovia, L., Piñero, D., Palacios, R., y Martínez-Romero, E. 1991. Genetic structure of a soil population of nonsymbiotic *Rhizobium leguminosarum*. *Appl. Environ. Microbiol.* 57:426-433.

Segovia, L., Young, J. P., y Martínez-Romero, E. 1993. Reclassification of American *Rhizobium leguminosarum* biovar *phaseoli* type I strains as *Rhizobium etli* sp. nov. *Int. J. Syst. Bacteriol.* 43:374-377.

Sessitsch, A., Hardarson, G., Akkermans, A. D. L., y de Vos, W. M. 1997a. Characterization of *Rhizobium etli* y other *Rhizobium* spp. that nodulate *Phaseolus vulgaris* L. in an Austrian soil. *Mol. Ecol.* 6:601-608.

Sessitsch, A., Ramírez-Saad, H., Hardarson, G., Akkermans, A. D., y de Vos, W. M. 1997b. Classification of Austrian rhizobia y the Mexican isolate FL27 obtained from *Phaseolus vulgaris* L. as *Rhizobium gallicum*. *Int. J. Syst. Bacteriol.* 47:1097-1101.

Silva, C., Eguiarte, L. E., y Souza, V. 1999. Reticulated y epidemic population genetic structure of *Rhizobium etli* biovar *phaseoli* in a traditionally managed locality in Mexico. *Mol. Ecol.* 8:277-287.

Silva, C., Vinuesa, P., Eguiarte, L. E., Martínez-Romero, E., y Souza, V. 2003. *Rhizobium etli* y *Rhizobium gallicum* nodulate common bean (*Phaseolus vulgaris*) in a traditionally managed milpa plot in Mexico:Population genetics y biogeographic implications. *Appl. Environ. Microbiol.* 69:884-893.

Silva, C., Vinuesa, P., Eguiarte, L. E., Souza, V., y Martínez-Romero, E. 2005. Evolutionary genetics y biogeographic structure of *Rhizobium gallicum sensu lato*, a widely distributed bacterial symbiont of diverse legumes. *Mol. Ecol.* 14: 4033-4050.

Singh, S. P., Gepts, P., y Debouck, D. G. 1991. Races of common bean (*Phaseolus vulgaris*, Fabaceae). *Econ. Bot.* 45:379-396.

Sivakumaran, S., Lockhart, P. J., y Jarvis, B. D. 1997. Identification of soil bacteria expressing a symbiotic plasmid from *Rhizobium leguminosarum* bv. *trofolii*. *Can. J. Microbiol.* 43:164-177.

Smith, M. W., Feng, D. F., y Doolittle, R. F. 1992. Evolution by acquisition:the case for horizontal gene transfers. *Trends Biochem. Sci.* 17:489-493.

Snyder, L., y Champness, W. 1997. Molecular genetics of bacteria, ASM Press, Washington, D. C.

Soberón-Chavez, G., y Nájera, R. 1988. Isolation from soil of *Rhizobium leguminosarum* lacking symbiotic information. *Can. J. Microbiol.* 35:464-468.

Souza, V., Bain, J., Silva, C., Bouchet, V., Valera, A., Márquez, E., y Eguiarte, L. E. 1997. Ethnomicrobiology:Do agricultural practices modify the population structure of the nitrogen fixing bacteria *Rhizobium etli* biovar *phaseoli*? *J. Ethnobiol.* 17:249-266.

Souza, V., Nguyen, T. T., Hudson, R. R., Piñero, D., y Lenski, R. E. 1992. Hierarchical analysis of linkage disequilibrium in *Rhizobium* populations:evidence for sex? *Proc. Natl. Acad. Sci. USA* 89:8389-8393.

Spaink, H. P., Kondorosi, A., y Hooykaas, P. J. J. 1998. The Rhizobiaceae, Kluwer Academic Publishers, Dordrecht.

Spratt, B. G., y Maiden, M. C. 1999. Bacterial population genetics, evolution y epidemiology. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* 354:701-710.

Stackebrandt, E., Frederiksen, W., Garrity, G. M., Grimont, P. A., Kampfer, P., Maiden, M. C., Nesme, X., Rossello-Mora, R., Swings, J., Truper, H. G., Vauterin, L., Ward, A. C., y Whitman, W. B. 2002. Report of the ad hoc committee for the re-evaluation of the species definition in bacteriology. *Int. J. Syst. Evol. Microbiol.* 52:1043-1047.

Stackebrandt, E., y Goebel, B. M. 1994. Taxonomic note:A place for ADN-ADN reassociation y 16S rARN sequence analysis in the present species definition in bacteriology. *Int. J. Syst. Bacteriol.* 44:846-849.

Sullivan, J. T., B. D. Eardly, P. van Berkum y C.W. Ronson. 1996. Four unnamed species of nonsymbiotic rhizobia isolated from the rhizosphere of *Lotus corniculatus*. *Appl. Environ. Microbiol.* 62:2818-2825.

Sullivan, J. T., H. N. Patrick, W. L. Lowther, D. B. Scott y C. W. Ronson. 1995. Nodulating strains of *Rhizobium loti* arise through chromosomal y symbiotic gene transfer in the environment. *Proc. Natl. Acad. Sci. USA* 92:8995-8999.

Sullivan, J. T., y Ronson, C. W. 1998. Evolution of rhizobia by acquisition of a 500-kb symbiosis isly that integrates into a phe-tARN gene. *Proc. Natl. Acad. Sci. USA* 95:5145-5149.

Sy, A., Giraud, E., Jourand, P., Garcia, N., Willems, A., de Lajudie, P., Prin, Y., Neyra, M., Gillis, M., Boivin-Masson, C., y Dreyfus, B. 2001. Methylotrophic *Methylobacterium* bacteria nodulate y fix nitrogen in symbiosis with legumes. *J. Bacteriol.* 183:214-220.

Tan, Z. Y., Kan, F. L., Peng, G. X., Wang, E. T., Reinhold-Hurek, B., y Chen, W. X. 2001. *Rhizobium yanglingense* sp. nov., isolated from arid y semi-arid regions in China. *Int. J. Syst. Evol. Microbiol.* 51:909-914.

Templeton, A. R. 1989. The meaning of species y speciation:A genetic perspective. En: Speciation y its consecuences. D. Otte y J. A. Endler (eds.). Sinauer, Sunderland, Massachusetts. Páginas 3-27.

Terefework, Z., Nick, G., Suomalainen, S., Paulin, L., y Lindstrom, K. 1998. Phylogeny of *Rhizobium galegae* with respect to other rhizobia y agrobacteria. *Int. J. Syst. Bacteriol.* 48 Pt 2:349-356.

Truchet, G., Rosenberg, C., Vasse, J., Julliot, J.-S., Camut, S., y J., D. 1984. Transfer of *Rhizobium meliloti* pSym genes into *Agrobacterium tumefaciens*:Host-specific nodulation by atypical infection. *J. Bacteriol.* 157:134-142.

Turner, S. L., y Young, J. P. 2000. The glutamine synthetases of rhizobia:phylogenetics y evolutionary implications. *Mol. Biol. Evol.* 17:309-319.

Ueda, T., Suga, Y., Yahiro, N., y Matsuguchi, T. 1995. Phylogeny of Sym plasmids of rhizobia by PCR-based sequencing of a *nodC* segment. *J. Bacteriol.* 177:468-472.

van Berkum, P., Beyene, D., Bao, G., Campbell, T. A., y Eardly, B. D. 1998. *Rhizobium mongolense* sp. nov. is one of three rhizobial genotypes identified which nodulate y form nitrogen-fixing symbioses with *Medicago ruthenica* [(L.) Ledebour]. *Int. J. Syst. Bacteriol.* 48:13-22.

van Berkum, P., y Eardly, B. D. 1998. Molecular evolutionary systematics of the *Rhizobiaceae*. Pages 1-24 in:The *Rhizobiaceae*:Molecular Biology of Plant-Associated Bacteria. H. P. Spaink, A. Kondorosi, y P. J. J. Hooykaas (eds.). Kluwer Academic Publishers, Dordrecht.

van Berkum, P., Terefework, Z., Paulin, L., Suomalainen, S., Lindstrom, K., y Eardly, B. D. 2003. Discordant phylogenies within the *rrn* loci of Rhizobia. *J. Bacteriol.* 185:2988-2998.

van Rhijn, P., y Vanderleyden, J. 1995. The *Rhizobium*-plant symbiosis. *Microbiol. Rev.* 59:124-142.

Vinuesa, P., León-Barrios, M., Silva, C., Willems, A., Jarabo-Lorenzo, A., Pérez-Galdona, R., Werner, D., y Martínez-Romero, E. 2005a. *Bradyrhizobium canariense* sp. nov., an acid-tolerant endosymbiont that nodulates endemic genistoid legumes (Papilionoideae:Genisteae) from the Canary Islands, along with *Bradyrhizobium japonicum* bv. *genistearum*, *Bradyrhizobium* genospecies α y *Bradyrhizobium* genospecies β. *Internation Journal of Syst. Evol. Microbiol.* 55:569-575.

Vinuesa, P., Rademaker, J. L. W., de Bruijn, F. J., y Werner, D. 1998. Genotypic characterization of *Bradyrhizobium* strains nodulating endemic woody legumes of the Canary Islands by PCR-restriction fragment length polymorphism analysis of genes encoding 16S rARN (16S rADN) y 16S-23S rADN intergenic spacers, repetitive extragenic palindromic PCR genomic fingerprinting y partial 16S rADN sequencing. *Appl. Environ. Microbiol.* 64:2096-2104.

Vinuesa, P., Rademaker, J. L. W., de Bruijn, F. J., y Werner, D. 1999. Characterization of *Bradyrhizobium* spp. strains by RFLP analysis of amplified 16S rADN y rADN intergenic spacer regions. Pages 275-279 in:Highlights on Nitrogen Fixation. E. Martínez y G. Hernández (eds.). Plenum Publishing Corporation, New York.

Vinuesa, P., y Silva, C. 2004. Species delineation y biogeography of symbiotic bacteria associated with cultivated y wild legumes. En: Biological Resources y Migration. D. Werner, ed. Springer Verlag, Berlin. Pp. 143-155.

Vinuesa, P., Silva, C., Lorite, M. J., Izaguirre-Mayoral, M. L., Bedmar, E. J., y Martínez-Romero, E. 2005b. Molecular systematics of rhizobia based on maximum likelihood y Bayesian phylogenies inferred from *rrs*, *atpD*, *recA* y *nifH* sequences, y their use in the classification of *Sesbania* microsymbionts from Venezuelan wetlands. *Syst. Appl. Microbiol.* 28: 702-716.

Vinuesa, P., Silva, C., Werner, D., y Martínez-Romero, E. 2005c. Population genetics y phylogenetic inference in bacterial molecular systematics:the roles of migration y recombination in *Bradyrhizobium* species cohesion y delineation. *Mol. Phylogenet. Evol.* 34:29-54.

Vulic, M., Dionisio, F., Taddei, F., y Radman, M. 1997. Molecular keys to speciation: ADN plymorphism y the control of genetic exchange in enterobacteria. *Proc. Natl. Acad. Sci. USA* 94:9763-9767.

Wang, E. T., Rogel, M. A., García-de los Santos, A., Martínez-Romero, J., Cevallos, M. A., y Martínez-Romero, E. 1999. *Rhizobium etli* bv. *mimosae*, a novel biovar isolated from *Mimosa affinis. Int. J. Syst. Bacteriol.* 49:1479-1491.

Ward, D. M. 1998. A natural species concept for prokaryotes. Curr. Opin. Microbiol. 1:271-277.

Weisburg, W. G., Barns, S. M., Pelletie, D. A., y Lane, D. J. 1991. 16S ribosomal ADN amplification for phylogenetic study. *J. Bacteriol.* 173:697-703.

Welch, R. A., Burland, V., Plunkett, G., 3rd, Redford, P., Roesch, P., Rasko, D., Buckles, E. L., Liou, S. R., Boutin, A., Hackett, J., Stroud, D., Mayhew, G. F., Rose, D. J., Zhou, S., Schwartz, D. C., PeARN, N. T., Mobley, H. L., Donnenberg, M. S., y Blattner, F. R. 2002. Extensive mosaic structure revealed by the complete genome sequence of uropathogenic *Escherichia coli*. *Proc. Natl. Acad. Sci. USA* 99:17020-17024.

Wernegreen, J., y Riley, M. A. 1999. Comparison of the evolutionary dynamics of symbiotic y housekeeping loci:A case for the genetic coherence of rhizobial lineages. *Mol. Biol. Evol.* 16:98-113.

Wernegreen, J. J., Harding, E. E., y Riley, M. A. 1997. *Rhizobium* gone native:Unexpected plasmid stability of indigenous *Rhizobium leguminosarum*. *Proc. Natl. Acad. Sci. USA* 94:5483-5488.

Willems, A., y Collins, M. D. 1993. Phylogenetic analysis of rhizobia y agrobacteria based on 16S rARN gene sequences. *Int. J. Syst. Bacteriol.* 43:305-313.

Woese, C. R. 2002. On the evolution of cells. Proc. Natl. Acad. Sci. USA 99:8742-8747.

Woese, C. R., Kandler, O., y Wheelis, M. L. 1990. Towards a natural system of organisms:proposal for the domains Archaea, Bacteria, y Eucarya. *Proc. Natl. Acad. Sci. USA* 87:4576-4579.

Wood, D. W., Setubal, J. C., Kaul, R., Monks, D. E., Kitajima, J. P., Okura, V. K., Zhou, Y., Chen, L., Wood, G. E., Almeida, N. F., Jr., Woo, L., Chen, Y., Paulsen, I. T., Eisen, J. A., Karp, P. D., Bovee, D., Sr., Chapman, P., Clendenning, J., Deatherage, G., Gillet, W., Grant, C., Kutyavin, T., Levy, R., Li, M. J., McClelland, E., Palmieri, A., Raymond, C., Rouse, G., Saenphimmachak, C., Wu, Z., Romero, P., Gordon, D., Zhang, S., Yoo, H., Tao, Y., Biddle, P., Jung, M., Krespan, W., Perry, M., Gordon-Kamm, B., Liao, L., Kim, S., Hendrick, C., Zhao, Z. Y., Dolan, M., Chumley, F., Tingey, S. V., Tomb, J. F., Gordon, M. P., Olson, M. V., y Nester, E. W. 2001. The genome of the natural genetic engineer *Agrobacterium tumefaciens* C58. *Science* 294:2317-2323.

Yap, W. H., Zhang, Z., y Wang, Y. 1999. Distinct types of rARN operons exist in the genome of the actinomycete *Thermomonospora chromogena* y evidence for horizontal transfer of an entire rARN operon. *J. Bacteriol.* 181:5201-5209.

Young, J. P. W. 1993. Molecular phylogeny of rhizobia y their relatives. En: New horizons in nitrogen fixation. P. R., J. Mora, y W. E. Newton, (eds.). Kluwer Academic Publishers, London. Páginas 587-592.

Young, J. P. W., y Haukka, K. 1996. Diversity y phylogeny of rhizobia. *New Phytol.* 133:87-94.

Young, J. P. W., y Johnston, A. W. B. 1989. The evolution of specificity in the legume-*Rhizobium* symbiosis. *Trends Ecol. Evol.* 4:341-349.

Young, J. P. W., y Wexler, M. 1988. Sym plasmid y chromosomal genotypes are correlated in field populations of *Rhizobium leguminosarum*. *J. Gen. Microbiol.* 134:2731-2739.

Capítulo 12

Ecología molecular en el estudio de comunidades bacterianas

Ana E. Escalante

Cuando se desea analizar a una especie o grupo biológico, es importante conocer el contexto ambiental y biológico en el que se encuentra. El contexto biológico se refiere a las especies con las que coexiste y en muchos casos de las que depende. El conjunto de especies que coexisten en un lugar y tiempo constituye la comunidad (Begon *et al.,* 1986).

En el caso de organismos macroscópicos es relativamente sencilla la caracterización y el conteo de las diferentes especies presentes en ella, pero en el caso de microorganismos esta caracterización se complica y es por ello que en los últimos años se han desarrollado diferentes estrategias, principalmente moleculares, para identificar las especies o grupos de microorganismos que constituyen la comunidad. Algunas de estas estrategias se explican y discuten en este capítulo y también se mencionan algunos ejemplos de este tipo de estudios. Pero antes de llegar a ello vale la pena establecer algunos antecedentes sobre diversidad bacteriana y cultivo de microorganismos que son relevantes para entender la importancia de desarrollar métodos finos que nos permitan descubrir la mayor parte de la diversidad que existe en las comunidades microbianas.

El problema: comunidades y diversidad bacteriana, microorganismos cultivables y no cultivables

La biomasa de microorganismos terrestres en la superficie de nuestro planeta ha sido estimada como igual a la de todas las plantas marinas y terrestres y tal vez sea el constituyente principal de la biomasa de la Tierra (Pedersen, 2000; Whitman *et al.,* 1998).

Los ambientes acuáticos parecen contar con menos diversidad que los suelos y los sedimentos. En los ambientes acuáticos extremos (por ejemplo, las pozas de desecación con 22% de sal) la diversidad genómica procarionte estimada por reasociación de ADN parece corresponder a sólo 7 genomas diferentes (Torsvic *et al.,* 1998; Øvreas *et al.,* 2001; Ritz *et al.,* 1997).

A pesar de que la diversidad genómica en el agua y el suelo puede ser distinta, el número máximo de individuos que integran el taxón dominante es el mismo, de aproximadamente 10^4 a 10^5 individuos por gramo o mililitro (Torsvik *et al.*, 2002). Este hallazgo sugiere que deben existir mecanismos que controlen la densidad de los taxa y que deben funcionar de manera más o menos similar en todos los ambientes, mientras que los mecanismos que controlan la abundancia total de la comunidad bacteriana funcionan de modos distintos en el agua y el suelo. Algunas ideas sobre los mecanismos que controlan la diversidad procarionte son las siguientes.

Interacciones tróficas

Hutchinson (1961) se preguntaba por qué hay tantas especies de fitoplancton en un medio aparentemente homogéneo, en donde todas las especies presentes parecen competir por el mismo tipo de nutrientes. Dentro de los procariontes los competidores pueden coexistir si existe algún mecanismo de pérdida selectiva de especies que evite que los competidores más exitosos secuestren todos los recursos (Bohannan y Lenski, 2000; Fuhrman 1999; Simek *et al.,* 1999).

Por ejemplo, existe depredación selectiva por tamaño llevada a cabo por protozoarios, lo que permite la coexistencia de bacterias y fitoplancton de diferentes tamaños y determina cómo se distribuye la biomasa en grupos funcionales. Otra interacción frecuente en el mundo de las comunidades microscópicas es la parasítica entre virus y bacterias, que es generalmente específica y permite la coexistencia de varios taxa de bacterias en una comunidad. Este tipo de interacción ha sido asociada con la especiación dentro

de grupos funcionales (Torsvic *et al.,* 2002), pues la transferencia horizontal mediada por fagos favorece la diversificación.

Perspectiva evolutiva

En trabajos sobre evolución experimental en bacterias se han observado altas tasas de diversificación; por ejemplo, a partir de una sola clona original se obtienen tres cepas ecológica y genéticamente distintas surgidas experimentalmente en el transcurso de 800 generaciones (Helling *et al.,* 1987; Rosenzweig *et al.,* 1994), lo que ha sugerido que la gran diversidad bacteriana que se detecta es simplemente una consecuencia de que las tasas de especiación son mayores que las tasas de extinción (Dykhuizen, 1998).

Esta idea deriva directamente de las observaciones experimentales en condiciones óptimas de crecimiento que muestran una rápida adaptación y partición del nicho que lleva a una diferenciación genética. Ésta eventualmente llevará a nuevas especies si se rompe el flujo génico entre las poblaciones diferenciadas.

Heterogeneidad espacio/temporal

La complejidad estructural del suelo y los sedimentos es importante para la diversificación en el nivel de poblaciones, porque permite que los recursos sean fraccionados y se creen nuevos nichos con los que se incrementa la especialización y división en especies ecológicas nuevas (Torsvic *et al.*, 2002).

La mayoría de las comunidades terrestres sufren perturbaciones intermitentes como escasez de alimento, sequía, congelamiento/descongelamiento o la actividad humana. Estas condiciones ambientales alteradas y los recursos que se liberan crean oportunidades para que nuevas especies se establezcan. Así, la perturbación asegurará que las comunidades incluyan una mezcla de los diferentes estadios de sucesión (Connell, 1978). Sin embargo, una perturbación fuerte y frecuente causará la desintegración de los microhabitats y la disrupción de los límites entre poblaciones, lo que permitirá que los recursos locales estén disponibles para una mayor proporción de la masa microbiana total (Torsvic *et al.*, 2002), es decir más individuos pero menos especies.

¿Cómo mejorar la recuperación en cultivo?

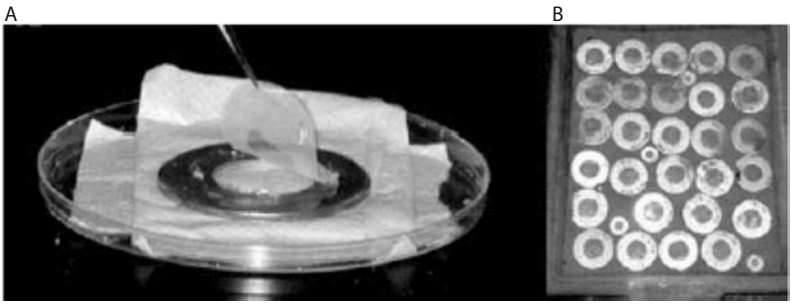
El interés en una mejor recuperación de microbios en cultivo se fundamenta principalmente en la posibilidad de obtener individuos sujetos de experimentación. El hecho de lograr tener en el laboratorio individuos microbianos abre la puerta a estudios sobre evolución, ecología y fisiología entre otros, lo que llevaría a un mayor y mejor conocimiento de estos organismos. El número de especies microbianas existentes ha sido estimado entre 10^5 y 10^6 (Tiedje 1995). Sin embargo, sólo se han aislado, descrito y/o caracterizado algunos miles. Esto se debe, en parte, a que sólo unos cuantos organismos de las muestras ambientales crecen en medios nutritivos en cajas de petri (Young, 1997; Barer y Harwood, 1999; Giovannoni y Rappé, 2000; Torsvic *et al.,* 1990).

Los intentos para mejorar la recuperación en cultivo han consistido en general en manipular cuidadosamente los medios de cultivo y aumentar el número de interacciones entre especies dentro de una misma caja (Kaeberlein *et al.,* 2002), pero se ha tenido poco éxito en ello y el problema del mundo no cultivable sigue siendo un reto importante de la microbiología (Young, 1997).

Kaeberlein *et al.* (2002) probaron si los organismos no cultivables crecerían si se les proveía de los componentes químicos de su ambiente nativo. Para permitir el acceso a estos componentes, colocaron a los microorganismos en cámaras de difusión (figura 1) y estas cámaras fueron incubadas en un acuario que simulaba el lugar nativo de estos organismos. Efectivamente lograron recuperar en cultivo cerca de 40% del inóculo original (se contó el número de colonias formadas en relación al número de células inoculadas), lo cual representa una recuperación en varios órdenes de magnitud mayor que la de los intentos tradicionales de cultivo. Además de esta extraordinaria recuperación "primaria" en cultivo, se intentaron aislar colonias de la comunidad cultivable y con este experimento se logró el resultado más significativo del trabajo. La mayoría de las colonias aisladas transferidas a cajas de Petri ($86\pm7\%$) no crecieron y las microcolonias que lograron crecer (14%) eran cultivos mixtos. Sólo unas pocas colonias grandes y que crecieron rápidamente fueron puras y representaron el $0.054\pm0.051\%$ del inóculo original, lo cual es consistente con lo que se ha reportado previamente en relación a la recuperación en cultivo.

Esta información lleva a conclusiones importantes sobre la ecología de los microorganismos, pues parece resaltar la importancia de posibles señales específicas originadas por los organismos vecinos que indiquen la presencia de un ambiente familiar. Esto significa que los microorganismos no crecerán en

Figura 1. Cámara de difusión para cultivo *in situ* de microorganismos ambientales. A) La cámara está formada por una "rondana" entre dos membranas de policarbonato con un poro de 0.03 mm; B) Las cámaras de crecimiento se incuban sobre la superficie de sedimento marino (tomada de Kaebelein *et al.*, 2002)



un ambiente extraño (sin sus vecinos) aunque existan los nutrientes apropiados (Kaeberlein *et al.,* 2002), lo que podría explicar por qué no es posible aislar a la mayoría de los microorganismos en medios artificiales como cultivos puros. Sin embargo habría que abrir la discusión sobre si realmente son los vecinos o hemos sido incapaces de determinar los micronutrientes necesarios para lograr cultivos puros de la mayoría de los microorganismos.

¿Qué hacer con los no cultivables? Métodos moleculares y otros métodos

A pesar de la posibilidad de aislar porcentajes mayores de microorganismos con estrategias novedosas de cultivo como la cámara de difusión de Kaeberlein, el porcentaje no cultivable sigue siendo un problema para la caracterización de la comunidad total. Así, se han desarrollado diferentes estrategias para caracterizar a las comunidades microbianas sin necesidad de cultivo y a continuación explicaremos los fundamentos de las más utilizadas.

PFLA (Phospholipid Fatty Acid Analysis)

En todos los microorganismos encontramos fosfolípidos exclusivos de las membranas celulares, que pueden ser degradadas fácilmente dejando libre al componente fosfolipídico. Los fosfolípidos de las membranas son indicadores

importantes de la biomasa microbiana activa y además se ha encontrado que existen lípidos únicos en grupos específicos de organismos, por lo que el análisis de fosfolípidos puede ser y ha sido usado para determinar la diversidad y abundancia microbianas (Hill *et al.,* 2000; Widmer *et al.,* 2001).

Esta estrategia requiere que la muestra ambiental sea analizada cuantitativamente por cromatografía de gases (GC) y espectrometría de masas (MS), lo que genera datos sobre la presencia y abundancia de diferentes fosfolípidos (Zelles y Bai 1993; Laczkó *et al.,* 1997). Los datos obtenidos sobre los ácidos grasos de las muestras pueden ser comparados con bases de datos públicas (Laczkó *et al.,* 1997), lo que permite la asociación de ciertos PFLAs con organismos o grupos específicos. Así, diferencias en los patrones de PFLA pueden ser interpretadas en relación con los cambios en la composición de las comunidades. Este método ha sido utilizado con éxito en estudios que demuestran cambios importantes en comunidades microbianas asociados a prácticas de manejo de suelos (Tunlid *et al.,* 1989; Bossio *et al.,* 1998).

A pesar de la utilidad de este método existen algunas limitaciones importantes (Haack *et al.,* 1994). La primera es que no se conocen los patrones de ácidos grasos para todos los organismos, por lo que los datos obtenidos de muchas muestras no pueden relacionarse con algún grupo. La segunda limitación es que los patrones de ácidos grasos son muy susceptibles a cambios que no se deben necesariamente a la composición de la comunidad, por lo que pueden darse patrones falsos que originen interpretaciones incorrectas. La tercera limitación es que bacterias y hongos producen diferentes cantidades de ácidos grasos dependiendo de las condiciones de crecimiento o estrés, por lo que a pesar de que los patrones de PFLA puedan correlacionarse con la presencia de algunos organismos, esto no significa necesariamente que esos patrones sean únicos para esos grupos en todas las condiciones.

Técnicas basadas en ácidos nucléicos

De entre todas las moléculas celulares probadas hasta ahora, los ácidos nucléicos han demostrado ser las moléculas más útiles, pues han contribuido con una nueva perspectiva al entendimiento de la estructura de las comunidades microbianas. Una de las principales ventajas asociadas a los métodos con ácidos nucleicos es que puede analizarse a la comunidad microbiana en su totalidad, incluyendo aquellos organismos que no han podido ser cultivados en el laboratorio. Es por ello que estos métodos se han vuelto cada vez más

importantes en la ecología microbiana (Pickup 1991; Stackebrant *et al.,* 1993; Amann *et al.,* 1995; Holben y Harri, 1995).

Aún los análisis de baja resolución y amplia escala, como la reasociación de ADN, permiten la determinación de la diversidad genética total de la comunidad bacteriana (Torsvik *et al.,* 1996). Electroforesis de productos de PCR (véase el artículo 17 de este libro) de genes de rARN en geles desnaturalizantes de gradiente (DGGE) permiten tener mayor resolución y proveen de información sobre cambios en el grueso de la estructura de la comunidad (Muyzer *et al.,* 1993). Cuando se combinan análisis de genes de rARN en DGGE con hibridación usando sondas filogenéticas o secuenciando se logra obtener la afiliación filogenética de los miembros dominantes en número que se encuentran en la comunidad (Øvreås *et al.,* 1997). La hibridación fluorescente *in situ* (FISH) de células microbianas con sondas filogenéticas provee de información sobre la composición taxonómica de la comunidad (Hahn *et al.,* 1992; Amann *et al.,* 1995). Finalmente, la clonación de productos de PCR de genes de rARN de la comunidad arroja información específica de la fracción no cultivable. Esta estrategia también permite la comparación de las estructuras de la fracción cultivable con la no cultivable (Stackebrandt y Rainey, 1995).

A continuación ampliaremos la descripción de las estrategias mencionadas en el párrafo anterior.

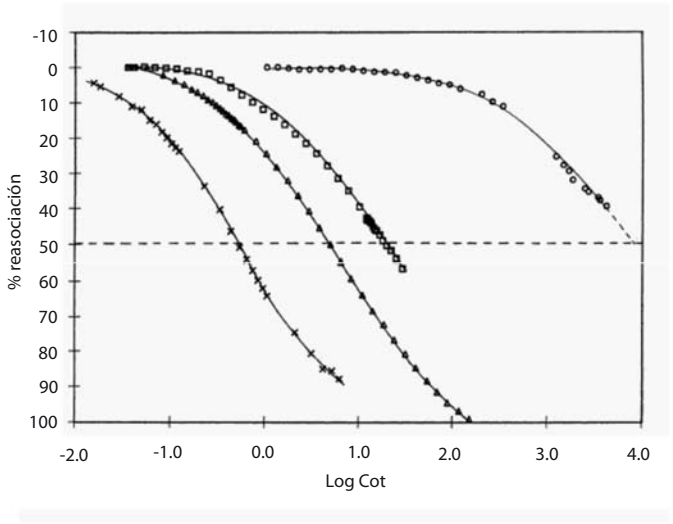
Cinética de reasociación de ADN

El ADN de una comunidad microbiana es una mezcla de genomas de diferentes especies presentes en distinta proporción. Una curva de reasociación de ADN nos da una idea de qué tan compleja es esa mezcla. La estrategia consiste en desnaturalizar con calor la mezcla total de ADN y mediante la lectura de un espectrofotómetro (ya que la monohebra de ADN y el ADN de doble cadena tienen picos de absorbancia a diferentes longitudes de onda; ver Torsvik *et al.,* 1995) se puede generar una cinética de reasociación de ADN. Esta cinética de reasociación es del tipo de segundo orden y cualquiera de este tipo de curvas tiene una constante de reasociación $C_0t_{1/2}$, donde C_0 es la concentración molar de nucléotidos en un ADN monohebra al inicio de la reasociación y $t_{1/2}$ es el tiempo en segundos en el que se da el 50% de la reasociación. Bajo condiciones definidas, $C_0t_{1/2}$ es proporcional a la complejidad del ADN (heterogeneidad; Britten y Kohne, 1968).

Estos valores son siempre relativos a la cinética de reasociación de ADN de una sola especie (*e.g.* ADN de *Escherichia coli*), que se considera la curva

ideal y una curva de una comunidad siempre tendrá pendientes menores que la ideal (figura 2). Aunque en estos experimentos $C_0t_{1/2}$ no tiene un significado preciso, este valor ha sido usado de modo similar a un índice de diversidad de especies (Torsvik *et al.,* 1995).

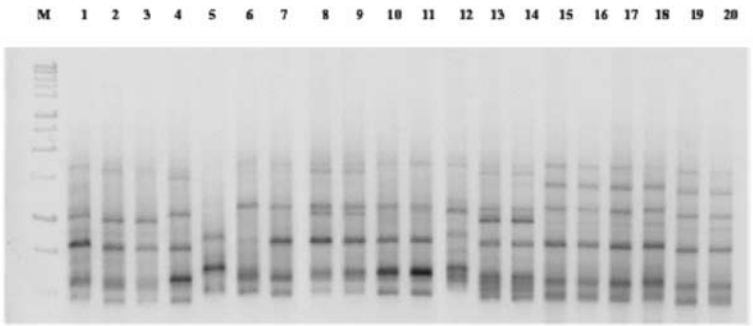
Figura 2. Cinética de reasociación de ADN de la fracción bacteriana de sedimento marino (o), mesocosmos con sedimento de granja de peces (□), mesocosmos con sedimento de granja de peces con 100ppm de ácido oxolínico (Δ), y *E. coli* (x) (tomado de Torsvik *et al.*, 1998)



RAPDs (Random Amplified Polymorphic DNA)

Dentro de los métodos de huellas de ADN o *fingerprint* basados en PCR (véanse los capítulos 17 y 18 de este libro) quizás el de RAPDs sea el más ampliamente usado. El PCR de RAPDs utiliza oligonucléotidos de 10 a 12 pares de bases para amplificar segmentos al azar de ADN. Este método funciona porque sólo un pequeño número de fragmentos (5 a 10) se amplificarán por estar dentro de un rango de longitud, usualmente menor a 3 o 4 kilobases, que puede ser fácilmente amplificado por PCR (Hoelzel y Green, 1998). Cuando el ADN que se encuentra entre las secuencias amplificadas consta de repeticiones en *tandem* y hay variaciones en su longitud, el fragmento amplificado será de longitud variable o polimórfico en la población o comunidad (figura 3).

Figura 3. Perfil de RAPD de la planta medicinal *Podophyllum peltatum*. Cada línea representa una población; la línea M corresponde al marcador de peso molecular de 1 kb (tomado de Lata *et al*., 2002)



Esta estrategia puede servir para dar un primer vistazo a la composición de distintas comunidades; sin embargo debemos mencionar que esta técnica es poco reproducible debido a que cualquier mínima variación en las condiciones de PCR o del ADN original puede generar cambios en los patrones, por lo que es recomendable utilizar técnicas adicionales para robustecer los datos.

Técnicas relacionadas con la diversidad de genes de rARN (SSU)

La determinación de secuencias o patrones del gen que codifica para la subunidad pequeña de los ribosomas (16S para procariontes y 5S o 18S para eucariontes; Ward *et al*., 1992), es una de las técnicas más útiles y ampliamente utilizadas.

Esta subunidad pequeña (SSU) de rARN es adecuada para este tipo de estudios por varias razones. La primera es que se encuentra en los tres dominios de la vida conocidas: Bacteria, Archaea y Eucarya (Woese *et al*., 1990). La segunda es que son moléculas que tienen regiones altamente conservadas y regiones de secuencia variable (Woese, 1987) y gracias a las tasas de evolución diferentes en cada tipo de región, se pueden establecer relaciones en diferentes niveles jerárquicos en un análisis comparativo de secuencias (Hill *et al.,* 2000). La tercera razón es que el rADN puede amplificarse fácilmente por PCR y ser secuenciado. La cuarta es que los microorganismos no culti-

vables pueden ser estudiados y caracterizados sin cultivarse al obtener por PCR su SSU rADN.

Varios estudios han demostrado que más del 90% de los microorganismos que se observan microscópicamente *in situ*, pueden ser extraídos y analizados (Steffan y Atlas 1998; Steffan *et al*., 1988; Tsai y Olsen, 1992; More *et al*., 1994; Zhou *et al*., 1996; Porteus *et al.,* 1997), en comparación con menos del 0.1% que pueden cultivarse (Hill *et al*., 2000). Una vez amplificado el gen de rARN por PCR es posible analizar su diversidad en la comunidad para tener una idea de la complejidad de la misma siguiendo diferentes estrategias que se explicarán a continuación.

DGGE/TGGE (Denaturing Gradient/ Temperature Gradient Gel Electrophoresis)

Esta técnica permite separar mezclas de productos del PCR que son de la misma longitud pero que difieren en secuencia. El poder de separación de este tipo de electroforesis se basa en que el comportamiento de desnaturalización del ADN está determinado de manera importante por la secuencia de nucléotidos. Al correr el ADN en el gel, la molécula se mantiene como doble cadena hasta que alcanza la concentración o temperatura desnaturalizante, reduciendo su movilidad en el gel (Hill *et al*., 2000).

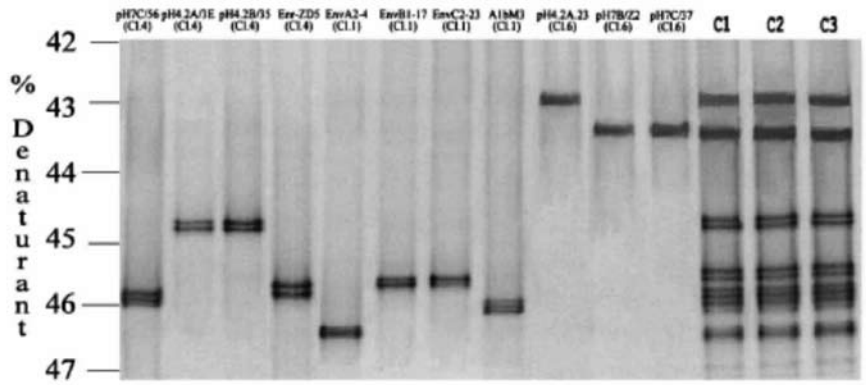
En teoría, cualquier gen de rARN que se encuentre en el ADN total de la comunidad puede ser amplificado por PCR y resuelto en un gel DGGE/TGGE, ya que en este tipo de análisis cada banda en el gel es considerada como una especie distinta en la comunidad y la intensidad de las bandas es tomada como un reflejo de la abundancia de esa secuencia en la comunidad (figura 4).

Existen algunas consideraciones importantes sobre esta técnica y tienen que ver con la interpretación de los datos obtenidos. En primer lugar es posible que diferentes especies generen una misma banda y que los estimados de diversidad no sean del todo estrictos, por lo que lo ideal es hacer estimaciones de diversidad relativa entre comunidades y considerar filotipos en lugar de especies.

T-RFLPs (Terminal Restriction Fragment Length Polymorphism)

El análisis de fragmentos de restricción terminales es actualmente uno de los métodos más poderosos dentro del campo de la ecología microbiana para comparar rápidamente la diversidad de secuencias de ADN bacteriano amplificado por PCR de muestras ambientales (Marsh, 1999; Tiedje *et al*., 1999).

Figura 4. Análisis de DGGE de la variación dentro y entre grupos de bacterias oxidadoras de amonio. El análisis fue hecho para secuencias de 11 clonas del gen 16S rARN, de las cuales 4 muestran afinidad por *Nitrospira* (carriles 1 al 4), 3 muestran afinidad con *Nitrosospira* (5 a 8) y 4 muestran afinidad con *Nitrosomonas* (8 a 11). . C1, C2 y C3 representan tres muestras separadas de PCR utilizando una combinación equimolar de las 11 clonas como templado



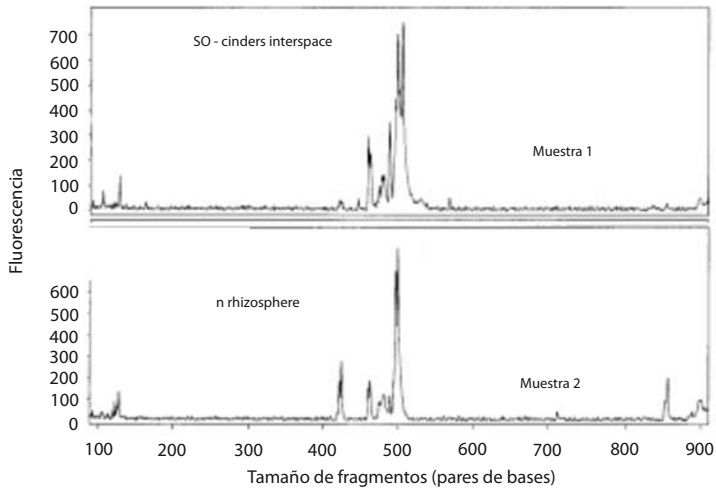
El método se basa en la variación en la posición de sitios de restricción entre las secuencias y en la determinación de la longitud de fragmentos terminales de restricción (TRFs) marcados con fluorescencia por medio de electroforesis en geles de alta resolución en secuenciadores de ADN (Avanis-Aghahani *et al.*, 1994; Liu *et al.*, 1997). El resultado es una distribución de abundancia de fragmentos de diferentes tamaños (figura 5).

Dentro de las características más importantes de este método está la gran velocidad de análisis de las muestras, lo que permite hacer réplicas de los experimentos y con ello lograr análisis estadísticos (Dunbar *et al.*, 2001). El método de TRF puede ser usado para identificar diferencias entre comunidades, comparar la riqueza relativa de filotipos y estructura de comunidades, y para identificar organismos específicos en una comunidad (Dunbar *et al.,* 2001) a partir de la clonación y secuenciación de fragmentos específicos de 16S rARN.

Sobre las limitaciones del método podemos decir que, a pesar de que es un método rápido de análisis comparativo de comunidades, tiene baja resolución ya que existe cierta probabilidad de que varios filotipos se encuentren representados por un solo tamaño de fragmento y representen un solo pico en la distribución, lo que puede llevar a subestimar la riqueza de especies, que

puede ser calculada por otros métodos con mayor resolución filogenética, aunque a mayor costo.

Figura 5. Patrones de T-RFLPs con la enzima Rsa I en 16S rADN amplificado directamente de muestras de suelo (tomado de Dunbar *et al.*, 2000)



Esta situación fue puesta a prueba por Dunbar *et al.,* (2001) al comparar los estimados de riqueza obtenidos de una librería de clonas de 16S y de TR-FLPs obtenidos de la misma muestra con la que se construyó la librería. La conclusión de Dunbar *et al.,* es que los patrones de TRFLPs son incapaces de proporcionar información confiable sobre el total de la riqueza relativa de los filotipos. Sin embargo, el método es muy efectivo en descubrir similitudes entre comunidades y tiene alta sensibilidad (detecta genomas que se encuentran en concentraciones tan bajas como el 1% del total de la comunidad). Es por ello que se recomienda como útil para análisis rápidos de muestras de campo que quieran compararse en cuanto a su composición y parece ser una nueva herramienta para investigaciones de ecología microbiana.

ARDRA (Amplified Ribosomal DNA Restriction Analysis)

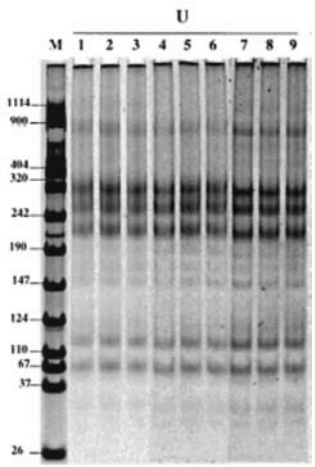
Es una técnica que consiste en amplificar por PCR el gen de 16S rARN de la comunidad total, aislar una a una las diferentes copias que se encuentran

en la mezcla y posteriormente someter a cada una de las copias a digestiones enzimáticas (restricción); también puede digerirse la muestra total (sin separar las copias; Martin-Laurent *et al.*, 2001).

La manera más común de separar las diferentes copias del 16S rARN (o cualquier otro gen) de la comunidad es clonando los diferentes fragmentos en cepas industriales de *Escherichia coli* transformable. Esto se hace al mezclar el PCR total con bacterias susceptibles a adquirir ADN extraño (competentes), por lo que cada bacteria adquirirá una sola copia del gen y así quedarán separadas las diferentes copias de la mezcla total. Las bacterias que contienen los fragmentos se crecen en cajas de Petri y posteriormente se aísla su ADN junto con el gen de interés que se puede separar del ADN bacteriano (véase los capítulos 17 y 18 de este libro). Cada gen obtenido de esta manera se somete a una digestión enzimática que genera un patrón de bandas en electroforesis en geles de agarosa (figura 6). Los patrones de restricción permiten identificar diferentes grupos (que no representan necesariamente especies diferentes) o identificar comunidades distintas.

Si el ARDRA se hace con copias separadas del gen 16S rARN, el análisis de los diferentes patrones permite hacer estimaciones de diversidad y/o elegir representantes de los patrones distintos para obtener las secuencias e identificar los grupos taxonómicos específicos a los que pertenecen o con los que se

Figura 6. Patrón de ARDRA de productos de PCR del gen 16S rARN (primers 27 y 1492) digeridos con RsaI de muestras ambientales de suelo. La línea M representa el marcador de peso molecular (tomada de Martin-Laurent *et al.*, 2001)



relacionan. El problema es que el número de clonas que se debe obtener por comunidad para determinar su diversidad y abundancia tiene un costo final similar al de la secuenciación de la comunidad.

Secuencias

Cualquiera de las estrategias de ácidos nucleicos mencionadas pueden concluir con la secuenciación de una muestra de las copias del gen 16S rARN. Sin embargo, se pueden evitar los análisis antes mencionados y hacer el PCR del ADN total e inmediatamente después secuenciar todas las copias con clonación previa. El análisis de secuencias en comunidades microbianas requiere siempre separar las diferentes copias de la muestra, ya sea por clonación o cortando las diferentes bandas de geles TGGE.

Obtener la secuencia de 16S rARN de los diferentes organismos de la muestra tiene varias ventajas. Una de ellas es que se construye una librería de clonas que mantiene las copias separadas y listas para otros análisis en el futuro. Otra ventaja y quizás la más importante es que la obtención de secuencias en librerías de clonas es la única manera de obtener estimados robustos sobre la diversidad de comunidades complejas (Dunbar *et al.*, 2000).

La principal limitante hasta ahora para la implementación general de análisis de comunidades con secuencias son sus costos económicos y de tiempo, ya que cada una de las secuencias debe ser "curada" confirmando que cada pico del electrofenograma sea el correcto. Las secuencias deben después ser sometidas a Blast con el GenBank y alineadas considerando la estructura terciaria del 16S. Afortunadamente estos costos se han ido abatiendo con el aumento de su demanda y el desarrollo de nuevas técnicas de secuenciación y clonación.

Técnicas relacionadas con la expresión de genes (rARN)

Recientemente varios análisis se han enfocado en la caracterización de comunidades microbianas del suelo basándose en la expresión del rARN en contraste con los genes que codifican para el ribosoma: el rADN (Felske y Akkermans, 1998b; Hahn *et al.,* 1990; Moran *et al.*, 1993; Felske *et al.*, 1996; Purdy *et al.*, 1996; Duarte *et al.*, 1998).

Al igual que el rADN, el rARN tiene regiones variables y regiones altamente conservadas que permiten la discriminación de taxa en diferentes niveles taxonómicos. Además, el uso de rARN directo ofrece varias ventajas sobre el uso de técnicas con rADN: debido a que los ribosomas son el sitio de síntesis

de proteínas, el contenido de rARN está directamente correlacionado con la actividad metabólica (Wagner, 1994), lo que significa que una alta proporción de las secuencias detectadas en las muestras corresponderá a microorganismos metabólicamente activos y en crecimiento (Felske *et al.*, 1996; Felske *et al.*, 1997; Felske y Akkermans, 1998a).

Además, debido a que las secuencias de rARN normalmente se encuentran en las células en más copias que las secuencias de rADN, teóricamente deben ser más fáciles de detectar. Al igual que con el rADN, cuando los amplicones de rARN se separan en geles DGGE/TGGE, el patrón de bandas sirve como una huella digital de la comunidad microbiana. Si se asume que no hay sesgo en la amplificación, la intensidad de las bandas indica la abundancia en la comunidad de la secuencia de rARN correspondiente (Felske *et al.*, 1998b).

Sin embargo, un factor importante que complica las interpretaciones sobre la abundancia de un taxon es que el número de operones de rARN dentro de un genoma puede variar entre los grupos taxonómicos (Rosado *et al.*, 1997), por lo que es posible que exista heterogeneidad en las secuencias de rARN dentro de un mismo grupo. En consecuencia no puede asumirse que cada producto de PCR en el gel corresponde a un organismo diferente, y un mismo organismo puede estar representado por varias secuencias distintas (Hill *et al.*, 2000). Por otro lado, la intensidad de la banda puede deberse a muchas copias de una misma secuencia en el mismo organismo o a que existen muchas células del organismo que posee la secuencia. Una manera de distinguir entre estas dos posibilidades es mediante hibridización *in situ* (ver Binder y Liu, 1998). Otra manera es correr el mismo gel con productos de PCR de rARN y rADN y comparar la intensidad de las bandas para determinar si ésta se encuentra relacionada con el número de copias del gen (Felske *et al.*, 1997).

A pesar de que esta técnica puede ser muy útil para estudiar comunidades microbianas, existen limitaciones técnicas en ella debido a lo lábil que es el ARN por lo que es necesario tener mucho cuidado en el almacenamiento y procesamiento de las muestras ambientales, ya que cualquier cambio mínimo en las condiciones ambientales originales puede alterar la actividad metabólica de los microbios (van Winzingerode *et al.*, 1997); esto puede ser evitado con el procesamiento inmediato o congelamiento de las muestras. Otra limitación existe en relación a la eficiencia de extracción (Moran *et al.*, 1993) y amplificación de rARN (Zheng *et al.*, 1996; Suzuki y Giovannoni, 1996; van Winzingerode *et al.*, 1997).

Existe la posibilidad de que la amplificación sea sesgada por una mayor abundancia de algunas secuencias en el ADN original o por homología entre ciertas secuencias y los oligonucleótidos del PCR. La mayoría de los estimados

son en procariontes y aunque en teoría debería poder analizarse la comunidad eucarionte de rARN de la misma manera que procarionte, los ribosomas eucariontes parecen ser más complejos en cuanto a los genes que los codifican y su regulación, además de que las bases de datos públicas con secuencias de rARN casi no tienen representantes eucariontes, por lo que identificar especies eucariontes con las secuencias de su rARN es problemático (Hill *et al.*, 2000).

FISH (Fluorescent in situ hybridization)

Este método ha sido usado principalmente con comunidades procariontes y permite la identificación y cuantificación directas de grupos taxonómicos generales o específicos dentro de su ambiente natural (Amann *et al.*, 1995; Assmus *et al.*, 1995; MacNaughton *et al.*, 1996; Kenzaka *et al.*, 1998).

Con FISH todas las células son fijadas y su 16S o 23S rARN se hibridiza con sondas de oligonucleótidos taxón específicos marcados con fluorescencia. Posteriormente, las células que hibridaron y quedaron marcadas fluorescentemente pueden verse en un microscopio óptico de fluorescencia. Con este método y gracias a que las células completas se hibridan, se evitan artefactos generados por sesgos en la extracción de ADN, amplificación por PCR y clonación (Ludwing *et al.*, 1997; Wallner *et al.*, 1997; Felske *et al.*, 1998a).

FISH puede usarse para visualizar organismos no cultivados, y es útil para estudiar la distribución ecológica de los organismos en hábitats diversos (Ludwig *et al.*, 1997; Zarda *et al.*, 1997; Wullings *et al.*, 1998). Tal vez el aspecto en el que se debe tener más cuidado al usar FISH es el diseño de la secuencia de las sondas de hibridación, ya que de ello depende la caracterización de la comunidad. Es necesario hacer una selección cuidadosa de las secuencias de los organismos representativos del taxón o taxa que se desea visualizar, pues la sonda será tan buena como lo sea esta selección (Amman *et al.*, 1995). Errores en este diseño pueden ocasionar que no puedan visualizarse organismos no cultivables que pertenezcan al grupo de interés o que se den hibridaciones cruzadas con organismos de otros grupos (Hahn *et al.*, 1992; MacNaughton *et al.*, 1996; Felske *et al.*, 1998a).

Métodos estadísticos para el análisis de las comunidades bacterianas

El análisis de cualquier comunidad involucra la determinación estadística de la diversidad a partir de la abundancia relativa de las especies presentes. Es por esto que la estimación de la diversidad en las comunidades bacterianas

es actualmente uno de los retos y objetivos más ampliamente perseguidos en ecología microbiana. Sin embargo, estimar la diversidad entraña dos problemas: determinar el tamaño de muestra necesario para tener representada a la comunidad y elegir el estimador de la diversidad más adecuado. El problema del tamaño de muestra no es único para las comunidades bacterianas. Los ecólogos del mundo macroscópico se han enfrentado al mismo problema y han diseñado herramientas para lidiar con el muestreo (Colwell y Coddington, 1994; Heck *et al.*, 1975; Magurran, 1995), que en la actualidad intentan aplicarse al mundo microscópico.

Hughes *et al.* (2001) publicaron una revisión interesante sobre este problema, y evaluaron el tamaño de muestra que en general se requiere para tener un estimado útil de la diversidad de las comunidades. Además de esto evaluaron la utilidad de varios estimadores de diversidad para el caso de comunidades bacterianas. Este apartado se centrará en estos puntos.

El tamaño de muestra

Curvas de acumulación de especies

La relación entre el número de tipos (OTUs = especies, filotipos, etc.; véase el capítulo 4 de este libro)) observados y el esfuerzo del muestreo nos da información sobre el total de la diversidad de la comunidad que ha sido muestreada. Este patrón puede visualizarse graficando una curva de acumulación que consiste en relacionar el número acumulativo de OTUs observados contra el esfuerzo de muestreo. Este tipo de curvas revelan la calidad del muestreo, pues ya que todas las comunidades contienen un número finito de especies, si se continúan muestreando individuos las curvas eventualmente alcanzarán una asíntota en el valor real del número de OTUs de la comunidad (figura 7).

En la mayoría de las comunidades microbianas muestreadas no se alcanza la asíntota de las curvas debido a la enorme diversidad que existe. Sin embargo, aunque sería útil conocer su diversidad real, la mayor parte de las preguntas se refieren a cambios provocados por alteraciones ambientales bióticas o abióticas, por lo que las respuestas a estas preguntas requieren únicamente estimados de diversidades relativas (Hughes *et al.*, 2001). La aproximación de diversidad relativa para evaluar el efecto de variables ambientales ha sido usada tanto en casos de insectos con diversidades altas a una escala de kilómetros cuadrados, como en la de bacterias en la boca

humana (Price, 1975; Samways, 1994) y esta estrategia parece estar difundiéndose y ser útil.

Rarefacción

Esta aproximación ha sido adoptada recientemente por varios microbiólogos (Bills y Polishook, 1994; Dunbar *et al.*, 1999; Moyer *et al.*, 1998). La rarefacción compara la diversidad observada entre sitios, tratamientos o hábitats que no han sido muestreados equitativamente. Para hacer una comparación justa entre sitios con tamaños de muestra distintos se extraen submuestras de la muestra más grande y se calcula la riqueza esperada de especies en las submuestras basándose en la distribución de abundancias de la muestra mayor. El proceso se repite para submuestras de diferentes tamaños. Una curva de rarefacción muestra el cambio en el valor esperado de riqueza de especies de acuerdo al tamaño de la muestra (figura 7; Ricklefs y Miller, 2000).

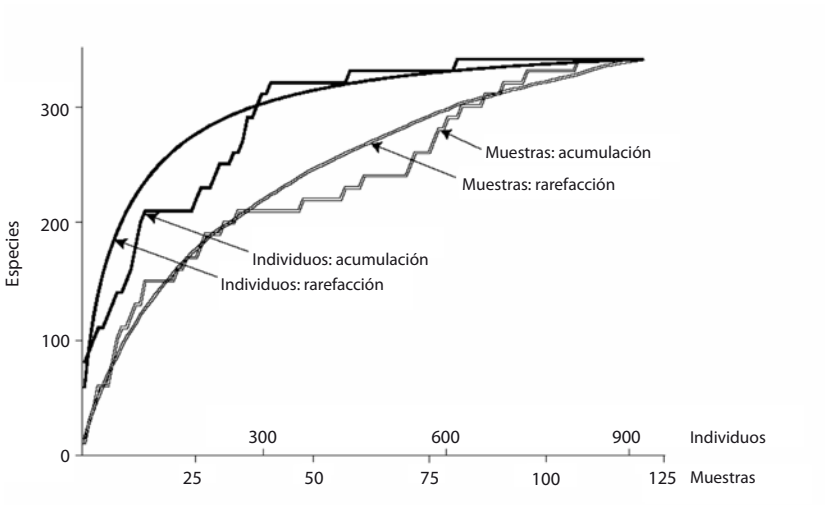
Estimadores de diversidad

Extrapolación a partir de curvas de acumulación

La mayoría de estos métodos usan la curva de acumulación observada para ajustarla a una forma funcional supuesta que modela el proceso de observar nuevas especies si aumenta el esfuerzo de muestreo (Hughes *et al.*, 2001). Estos modelos incluyen la ecuación de Michaelis-Menten (Clench, 1979; Raaijamakers, 1987) y la función exponencial negativa (Soberón y Llorente, 1993).

El beneficio de estudiar la diversidad con estos métodos es que una vez que las especies se han contado no es necesario contarlas de nuevo y el esfuerzo sólo debe enfocarse en encontrar especies raras nuevas. La desventaja es que para comunidades muy diversas sólo es posible detectar una pequeña fracción de las especies y es posible que varias curvas de acumulación tengan un buen ajuste pero que cada una prediga diferente asíntota (Soberón y Llorente, 1993). Esta aproximación de datos se utiliza para comunidades de macroorganismos relativamente bien muestreados por lo que no parece útil por el momento para comunidades bacterianas (Hughes *et al.*, 2001).

Figura 7. Las curvas de acumulación representan el ordenamiento único de individuos (línea sólida e irregular), o de muestras (línea hueca e irregular) conforme se van acumulando los datos. Las curvas de rarefacción representan las medias de re-muestreos repetidos de todos los individuos de la muestra (línea sólida y regular), o de todas las muestras (línea hueca y regular). Las curvas de rarefacción representan la expectativa estadística para las correspondientes curvas de acumulación. Las cuatro curvas están basadas en la cuenta de semillas del grupo de datos de Butler y Chazdon (1998), analizados por Colwell y Coddington (1994) y disponibles en línea con EstimateS (Colwell, 2000) (tomada de Gotelli y Colwell, 2001)



Estimadores paramétricos

Los métodos paramétricos estiman el número de especies no observadas en la comunidad al ajustar los datos del muestreo a modelos de abundancia relativa de especies. Estos modelos incluyen el modelo lognormal (Preston, 1948) y el modelo Poisson normal (Bulmer, 1974). Al ajustar los datos a estas distribuciones pueden obtenerse los parámetros de la curva, entre ellos la diversidad de la comunidad. Sin embargo, existen algunos impedimentos para usar estimadores paramétricos para cualquier comunidad. Uno de los más importantes es que se asume que se conoce la distribución de abundancia. A pesar de que muchas comunidades de macroorganismos parecen tener una distribución lognormal de abundancia de especies (DeVries *et al.,* 1997; Magurran, 1995; Sugihara,

1980), aún existe controversia sobre el modelo que se ajusta mejor (Harte *et al.,* 1999; Hughes, 1986). El problema es mucho mayor en las comunidades microbianas, ya que en ausencia de buenas bases de datos sobre diversidad y abundancia no es claro si alguno de estos modelos se ajusta a la distribución real. Otra limitación es que los estimadores paramétricos requieren grandes series de datos sobre abundancia para evaluar los parámetros de la distribución y en el caso de microorganismos las series de datos más grandes que existen incluyen sólo algunos cientos de individuos (Hughes *et al.,* 2001).

Estimadores no paramétricos

La diversidad de una comunidad depende de la riqueza (número de especies) y de la abundancia relativa de éstas (equitatividad o *evenness*; Begon *et al.,* 1986). Los estimadores no paramétricos que existen ponderan de diferente manera a las especies abundantes en relación con las especies raras. La medida más simple de diversidad que toma en cuenta tanto la abundancia como la riqueza de especies es el índice de diversidad de Simpson.

Este índice *D* se calcula a partir de la proporción de individuos de cada especie *i* que contribuyen al total de la muestra, donde el límite superior *S* se refiere al número total de especies de la comunidad (Begon *et al.,* 1986):

$$D=1/\Sigma p_i^2$$

Este índice está basado en la probabilidad de que si se toman dos individuos, estos sean de la misma especie (Brown y Bowman, 2001). Valores cercanos a 1 indican que sólo unas pocas especies predominan en la muestra.

Otro índice de diversidad muy utilizado es el denominado índice de Shannon-Weaver:

$$H=-\Sigma\, p_i(ln\ (p_i))$$

Este índice calcula la probabilidad de predecir la especie de un individuo tomado al azar de la comunidad. El índice va de 0 para comunidades con una sola especie hasta valores altos para comunidades con muchas especies, cada una con pocos individuos (Brown y Bowman, 2001).

Otros estimadores no paramétricos que parecen ser muy prometedores para la ecología microbiana han sido adaptados a partir de estadísticos de marcaje-liberación-recaptura (MRR) para estimar el tamaño de poblaciones

animales (Krebs, 1989; Seber, 1973). Los estimadores no paramétricos basados en métodos MRR consideran la proporción de especies que han sido observadas antes (recapturadas) en relación a aquellas que han sido observadas sólo una vez. En una comunidad muy diversa, la probabilidad de que las especies sean observadas más de una vez será baja y la mayoría de las especies serán representadas por un solo individuo. Los estimadores Chao1 y ACE (cobertura basada en la abundancia) usan la relación MRR para estimar la diversidad como riqueza total de especies al agregar un factor de corrección al número observado de especies (Chao, 1984; Chao *et al.,* 1993). Chao1 estima la riqueza total de especies como:

$$S_{Chao1} = S_{obs} + (n_1^2/2n_2)$$

En donde S_{Chao1} es la riqueza total de especies, S_{obs} es el número de especies observadas, n_1 es el número de especies observadas una vez y n_2 es el número de especies observadas dos veces. Chao (1984) hace notar que este índice es particularmente útil para series de datos que tengan poca abundancia en sus clases, como es el caso de los microbios (Hughes *et al.,* 2001). En cambio, ACE (Chao y Lee, 1992) incorpora datos de todas las especies con menos de 10 individuos en lugar de sólo vistas 1 o 2 veces:

$$S_{ACE} = S_{abund} + (S_{rare}/C_{ACE}) + (F_1/C_{ACE})$$

donde S_{ACE} es la riqueza total de especies, S_{abund} es el número de muestras abundantes (>10), S_{rare} es el número de muestras raras (≤10), C_{ACE} es la cobertura de la muestra (proporción de especies raras).

Tanto Chao1 como ACE subestiman la riqueza verdadera cuando los tamaños de muestra son pequeños (Colwell y Coddington, 1994). En el análisis de los estimadores de diversidad para comunidades bacterianas realizado por Hughes *et al.* (2001), se calcularon y compararon diversidades microbianas en variedad de ambientes y encontraron que aunque los estimadores dependen del tamaño de la muestra, la mayoría de los estimadores de riqueza se estabilizan con los tamaños de muestra que, en general, se tienen para microbios (200 a 1000 clonas).

Es importante señalar que todas las estimaciones estadísticas tienen sus limitaciones y una de las más importantes y frecuentes es que no existe rigor en la definición de los OTUs (Staley, 1997), por lo que estudios distintos que calculen los mismos estimadores en general no son del todo comparables.

De igual manera, la mayoría de estas aproximaciones requieren datos sobre frecuencias relativas de los diferentes OTUs y muchos estudios han revelado que los análisis genéticos de diversidad microbiana van acompañados de sesgos de muestreo, como los ya mencionados durante la extracción de ADN y PCR. Sin embargo, el hecho de que la mayoría de las preguntas sobre estructura y función de las comunidades requieran comparaciones relativas sobrepasa muchos de los problemas sobre definición de especies y sesgos de muestreo (Hughes *et al.,* 2001). Mientras que la unidad de medida sea definida y constante, la diversidad entre sitios o tratamientos puede ser comparada. Asimismo, para minimizar el efecto del sesgo en el muestreo se pueden usar varias técnicas o genes que hagan más robustas las comparaciones (Nübel *et al.,* 1999b).

Revisión de estudios empíricos, crítica y perspectivas

Como ya hemos revisado a lo largo de este capítulo, el estudio de las comunidades microbianas inició con la investigación de los organismos que era posible cultivar en el laboratorio, pero esta estrategia ha resultado ser poco eficaz, ya que, como se mencionó, se ha sugerido que sólo el 0.1% de la comunidad total se logra aislar en cultivo (Hill *et al.,* 2000). A pesar de este bajo porcentaje recuperado y estudiado, la diversidad total encontrada es muy alta (alrededor de 4500 especies caracterizadas a la fecha, ver Torsvic *et al.,* 2002; véanse las críticas al concepto de especie bacteriana en los capítulo 10 y 11 de este libro).

Los estudios con comunidades cultivables empezaron a darnos una idea de la cantidad de grupos o especies microbianas que pueden existir y de la diversidad y riqueza de especies que puede haber en diferentes habitats. Los típicos trabajos de caracterización de comunidades cultivables se basaron en los diferentes tipos de metabolismo que existían en la comunidad (Garland y Lehman, 1999) y más recientemente este mismo enfoque ha sido industrializado en un producto conocido como BIOLOG™ . Este producto está basado en el uso de un juego de 95 fuentes diferentes de carbono que han sido descritas previamente (Garland y Mills, 1991). Sin embargo, existen varias consideraciones importantes en el uso de este método para el análisis de comunidades. Una de las más importantes es que los substratos utilizados por el sistema pueden no ser los relevantes ecológicamente y muy probablemente no reflejen la diversidad de substratos que encontramos en el ambiente.

Además, mientras que los perfiles de la fisiología de las comunidades pueden dar información sobre la diversidad microbiana, este y cualquier otro método basado en cultivo tiene aún el gran sesgo de muestreo generado por la limitante de las estrategias actuales de cultivo.

En prácticamente todos estos estudios se han descubierto nuevos linajes o probado nuevas técnicas de estudio y análisis de comunidades microbianas. Este proceso también ha aportado nuevo conocimiento sobre la distribución de varios grupos microbianos. Tal es el caso de la presencia de miembros de Archaea en comunidades mésicas (Ueda *et al.,* 1995), a pesar de que se consideró por muchos años que los organismos de este grupo sólo se encontraban en ambientes extremos. Sin embargo, esta importante área de la ecología microbiana sigue en su fase descriptiva ya que la mayoría de los trabajos actuales buscan resolver preguntas sobre la diversidad y el efecto de los factores ambientales o humanos sobre ella, así como su determinación en diferentes, tratando también de determinar la utilidad y alcances de las diferentes estrategias de muestreo (Dunbar *et al.,* 1999, 2000, 2001; McCaig *et al.,* 1999; Nübel *et al.,* 1999a; Kroes *et al.,* 1999; Hoshino *et al.,* 2001; Wagner *et al.,* 2002; Zhou *et al.,* 2002). Entre las estrategias actualmente más utilizadas para determinar la diversidad microbiana se encuentran la construcción de librerías de clonas y FISH para identificar la distribución de las especies.

Todos estos esfuerzos están contribuyendo rápidamente a la generación de inmensas bases de datos sobre la abundancia y distribución de las especies microbianas. El carácter claramente descriptivo de los trabajos actuales sobre comunidades microbianas se debe a que es un campo que apenas empieza a desarrollarse, ya que las herramientas para su estudio son relativamente recientes. Sin embargo, este es un primer paso para cambiar la tendencia que permaneció por mucho tiempo en el estudio de microbios, en la que los microbiólogos pasaron décadas analizando organismos cultivables y caracterizándolos de acuerdo con su metabolismo. Ahora, gracias a los estudios sobre microorganismos no cultivables se sabe que la proporción de conocimiento que estos años de microbiología clásica han dejado es limitada, aunque valiosa. Sin embargo nos encontramos apenas al pie de la montaña. Como hemos visto, muchos grupos de investigadores se dedican básicamente a hacer listados de abundancia y diversidad de las comunidades, a probar nuevos métodos de obtención y análisis de datos y a muestrear la mayor cantidad de hábitats diferentes, lo que refleja una posición de recopilación de información. Daría la impresión de que esta labor sería interminable pero poco a poco nos vamos dando cuenta de que no es imposible, las herramientas moleculares

y tecnológicas dan la posibilidad de agilizar este proceso. Aún falta por dar el salto hacia interpretaciones que integren la información de estos listados, integración que esperamos nos dé cada día una idea más clara de cómo funciona el mundo microscópico para contar en el futuro con información que ayude a analizar ampliamente la ecología y evolución de esta diversidad en ambientes naturales.

Bibliografía

Amann, R., W. Ludwing y K.H.Scleifer. 1995. Phylogenetic identification and *in situ* detection of individual microbial cells without cultivation. *Microbiol. Rev.* 59:143-169.

Assmus, B., P. Hutzler, G. Kirchhof, R. Amman, J.R. Lawrence y A. Hartmann. 1995. *In situ* localization of *Azospirillum brasilense* in the rizhosphere of wheat with fluorescently labeled rRNA-targeted oligonucleotide probes and scanning confocal laser microscopy. *Appl. Environ. Microbiol.* 61:1013-1019.

Avaniss-Aghajani, E., K. Jones, D. Chapman y C. Brunk. 1994. A molecular technique for identification of bacteria using small subunit ribosomal RNA sequences. *BioTechniques* 17:144-149.

Barer, M.R. y C.R. Harwood. 1999. Bacterial viability and culturability. *Adv. Microb. Physiol.* 41:93-137.

Begon, M., J. Harper y C.R. Townsend. 1986. *Ecology - Individuals, Populations and Communities*. Blackwell Scientific Publications. Londres.

Bills, G.F. y J.D. Polishook. 1994. Abundance and diversity of microfungi in leaf-litter of a lowland rain-forest in Costa Rica. *Mycologia* 86:187-198.

Binder, B.J. y Y.C.Liu. 1998. Growth rate regulation of rRNA content of marine *Synechococcus* (cyanobacterium) strain. *Appl. Environ. Microbiol.* 64:3346-3351.

Bohannan, B.J.M. y R.E. Lenski. 2000. The relative importance of competition and predation varies with productivity in a model community. *Am. Nat.* 156:329-340.

Bossio, D.A., K.M. Scow, N. Gunapala y K.J. Graham. 1998. Determinants of soil microbial communities: effects of agricultural management season and soil type on phospholipid fatty acid profiles. *Microb. Ecol.* 36:1-12.

Britten, R.J. y D.E. Kohne. 1968. Repeated sequences in ADN. *Science*161:529-540.

Brown, M.V. y J.P. Bowman. 2001. A molecular phylogenetic survey of sea-ice microbial communities (SIMCO). *FEMS Microbiol. Ecol.* 35:267-275.

Bulmer, M.G. 1974. On fitting the Poisson lognormal distribution to species abun-

dance data. *Biometrics* 30:101-110.
Chao, A. 1984. Non-parametric estimation of the number of classes in a population. *Scand. J. Stat.* 11:265-270.
Chao, A. y Lee S.-M. 1992. Estimating the number of classes via sample coverage. *J. Am. Stat. Assoc.* 87:210-217.
Chao, A., M.-C. Ma y M.C.K. Yang. 1993. Stopping rules and estimation for recapture debugging with unequal failure rates. *Biometrics* 43:783-791.
Clench, H. 1979. How to make regional lists of butterflies: Some thoughts. *J. Lepid. Soc.* 33:216-231.
Colwell, R.K. y J.A. Coddington. 1994. Estimating terrestrial biodiversity through extrapolation. *Phil. Trans. R. Soc. London B* 345:101-118.
Connell, J.H. 1978. Diversity in tropical rain forests and coral reefs. *Science* 199:1302-1310.
DeVries, P.J., D. Murray y R. Lande. 1997. Species diversity in vertical, horizontal, and temporal dimensions of a fruit-feeding butterfly community in an Ecuatorian rainforest. *Biol. J. Linnean Soc.* 62:343-364.
Duarte, G.F., A.S. Rosado, L. Seldin, A.C. Keijzer-Wolters y J.D. van Elsas. 1998. Extraction of ribosomal RNA and genomic ADN from soil for studying the diversity of the indigenous bacterial community. *J.Microbiol. Meth.* 32:21-29.
Dunbar J., S. Takala, S.M. BRNAes, J.A. Davis y C.R. Kuske. 1999. Levels of bacterial community diversity in four arid soils compared by cultivation and 16S rRNA gene cloning. *Appl. Environ. Microbiol.* 65:1662-1669.
Dunbar, J., L.O. Ticknor y C.R. Kuske. 2000. Assessment of microbial diversity in four southwestern United States soils by 16S rRNA gene terminal restriction fragment analysis. *Appl. Environ. Microbiol.* 66(7):2943-2950.
Dunbar J., L.O. Ticknor y C.R. Kuske. 2001. Phylogenetic specificity and reproducibility and new method for analysis of terminal restriction fragment profiles of 16s rRNA genes from bacterial communities.
Dykhuizen, D.E. 1998. Santa Rosalia revisited: why are there so many species of bacteria? *Antoine van Leeuwenhoek* 73:25-33.
Felske, A., B. Engelen, U. Nubel y H. Backhaus. 1996. Direct ribosome isolation from soil to extract bacterial rRNA for community analysis. *Appl. Environ. Microbiol.* 62:4162-4167.
Felske, A., H. Rheims, A. Wolterink, E. Stakebrandt y A.D.L. Akkermans. 1997. Ribosome analysis reveals prominent activity of an uncultured member of the class Actinobacteria in grassland soils. *Microbiolology* 143:2983-2989.
Felske, A. y A.D.L. Akkermans. 1998a. Prominent occurrence of ribosomes from an uncultured bacterium of the Verrrucomicrobiales cluster in grassland soils. *Lett.*

Appl. Microbiol. 26:219-223.

Felske, A. y A.D.L. Akkermans. 1998b. Spatial homogeinity of abundant bacterial 16S rRNA molecules in grassland soils. *Microb. Ecol.* 36:31-36.

Felske, A., A.D.L. Akkermans y W. De Vos. 1998a. *In situ* detection of an uncultured predominant *Bacillus* in Dutch grassland soils. *Appl. Environ. Microbiol.* 64:4588-4590.

Felske, A., A.D.L. Akkermans y W. De Vos. 1998b. Quantification of 16S rRNAs in complex bacterial communities by multiple competitive reverse transcription-PCR in temperature gradient gel electrophoresis fingerprints. *Appl. Environ. Microbiol.* 64:4581-4587.

Fuhrman, J.A. 1999. Marine viruses and their biogeochemical and ecological effects. *Nature* (London) 399:541-548.

Garland, J.L. y R.M. Lehman. 1999. Dilution/extintion of community phenotypic characters to estimate relative structural diversity in mixed communities. *FEMS Microbiol. Ecol.* 30:333-343.

Garland, J.L. y A.L. Mills.1991. Classification and characterization of heterotrophic microbial communities on the basis of patterns of community-level sole-carbon source-utilization. *Appl. Environ. Microbiol.* 57:2351-2359.

Giovannoni, S.G. y M. Rappé. 2000. Evolution, diversity and molecular ecology of marine prokaryotes. En: D.L. Kirchman [ed.]. *Microbial Ecology of the Oceans*. Wiley-Liss, Nueva York.

Haack, S.K., H. Garchow, D.A. Odelson, L.J. Forney y M.J. Klug. 1994. Accuracy, reproducibility and interpretation of fatty acid methyl ester profiles of model bacterial communities. *Appl. Environ. Microbiol.* 60:2483-2493.

Hahn, D., R. Kester, M.J.C. Starrenburg y A.D.L. Akkermans.1990. Extraction of ribosomal RNA from soil for detection of *Frankia* with oligonucleotide probes. *Arch. Microbiol.* 154:329-335.

Hahn, D., R.I. Amann, W. Ludwing, A.D.L. Akkermans y K.H. Schelifer. 1992. Detection of microorganisms in soil after in situ hybridization with rRNA-targeted, fluorescently labelled oligonicleotides. *J. Gen. Microbiol.* 138:1307-1312.

Harte, J., A. Kinzig y J. Green. 1999. Self-similarity in the distribution and abundance of species. *Science* 284:334-336.

Heck, K.L., G.V. Belle y D. Simberloff. 1975. Explicit calculation of the rarefaction diversity measurement and the determination of sufficient sampling size. *Ecology* 56:1459-1461.

Helling, R.B., C. Vargas y J. Adams. 1987. Evolution of *Escherichia coli* during growth in a constant environment. *Genetics* 116:349-358.

Hill, G.T., N.A. Mitkowski, L. Aldrich-Wolfe, L.R. Emele, D.D. Jurkonie, A. Ficke,

S. Maldonado-Ramírez, S.T. Lynch y E.B. Nelson. 2000. Methods for assesing the composition of soil microbial communities. *Applied Soil Ecology* 15:25-36.

Hoelzel, A.R. y A. Green. 1998. PCR protocols and analysis by direct ADN sequencing and PCR-based ADN fingerprinting. En: A.L. Hoelzel [ed.]. *Molecular Genetic Analysis of Populations – A practical approach*. Oxford University Press, Nueva York. pp.

Holben, W.E. y D. Harris. 1995. ADN-based monitoring of total bacterial community structure in environmental samples. *Mol. Ecol.* 4:627-631.

Hoshino, T., N. Noda, S. Tsuneda, A. Hirata y Y. Inamori. 2001. Direct detection by in situ PCR of the amoA gene in biofilm resulting from a nitrogen removal process. *Appl. Environ. Microbiol.* 67(11):5261-5266.

Hughes, J.B., J.J. Hellmann, T.H. Rickets y J.M. Bohannan. 2001. Counting the uncountable: statistical approaches to estimating microbial diversity. *Appl. Environ. Microbiol.* 67:4399-4406.

Hughes, R.G. 1986. Theories and models of species abundance. *Am. Nat.* 128:897-899.

Hutchinson, G.E. 1961. The paradox of the plankton. *Am. Nat.* 95:137-147.

Kaeberlein, T., K. Lewis y S.S. Epstein. 2002. Isolating the "uncultivable" microorganisms in pure culture in a simulated natural environment. *Science* 296:1127-1129.

Kenzaka, T., N. Yamaguchi, K. Tani y M. Nasu. 1998. rRNA –targeted fluorescent *in situ* hybridization analysis of bacterial community structure in river water. *Microbiology* 144:2085-2093.

Krebs, C.J. 1989. *Ecological methodology*. Harper and Row, Nueva York.

Kroes, I., P.W. Lepp y D.A. Relman. 1999. Bacterial diversity within the human subgingival crevice. *Proc. Natl. Acad. Sci. USA* 96:14547-14552.

Laczkó, E., A. Rudaz y M. Aragno. 1997. Diversity of antropogenically influenced or distributed soil microbial communities. En: H. Insam y A. Rangger [eds.]. *Microbial Communities – Functional versus structural approaches*. Springer Verlag, Heidelberg.

Lata, H., R.M. Moraes, A. Douglas y B.E. Scheffler. 2002. Assessment of genetic diversity in *Podophyllum peltatum* by molecular markers. En: J. Janick y A. Whipkey (eds.). *Trends in new crops and new uses.* ASHS Press, Alexandria, VA. pp. 537–544.

Liu, W., T.L. Marsh, H. Cheng y L.J. Forney. 1997. Characterization of microbial diversity by determining terminal restriction fragment length polymorphisms of genes encoding 16S rRNA. *Appl. Environ. Microbiol.* 63:4516-4522.

López-García, P., F. Rodríguez-Valera, C. Pedrós-Alió y D. Moreira. 2001. Unex-

pected diversity of small eukaryotes in deep-sea Antarctic plankton. *Science* 409:603-607.

Ludwing, W., S.H. Bauer, M. Bauer, I. Held, G. Kirchhof, R. Chulze, I. Huber, S. Spring, A. Hartmann y K.H. Schleifer. 1997. Detection and *in situ* identification of representatives of a widely distributed new bacterial phylum. *FEMS Microbiol. Lett.* 153:181-190.

MacNaughton, S.J., T. Booth, T.M. Embley y A.G. O´Donnell. 1996. Physical stabilization and confocal microscopy of bacteria on roots using 16S rRNA targeted fluorescent-labeled oligonucleotide probes. *J. Microbiol. Meth.* 26:279-285.

Magurran, A.E. 1995. Ecological diversity and its measurement. Princeton University, Princeton, N.J.

Marsh, T.L. 1999. Terminal restriction fragment length polymorphism (T-RFLP): an emerging method for characterizing diversity among homologous populations of amplification products. *Curr. Opin. Microbiol.* 2:323-327.

Martin-Laurent, F., L. Philippot, S. Hallet, R. Chaussod, J.C. Germon y G. Soulas. 2001. ADN Extraction from soils: old bias for new microbial diversity analysis methods. *Appl. Environ. Microbiol.* 67(5):2354-2359.

McCaig, A.E., L. Glover y J.I. Prosser. 1999. Molecular analysis of bacterial community structure and diversity of unimproved and improved upland grass pastures. *Appl. Environ. Microbiol.* 65:1721-1730.

Moran, M.A., V.L. Torsvik, T. Torsvik y R.E. Hodson. 1993. Direct extraction and purification of rRNA for ecological studies. *Appl. Environ. Microbiol.* 59:915-918.

More, M.I., J.B. Herrick, M.C. Silva, W.C. Ghiorse y E.L. Madsen. 1994. Quantitative cell lysis of indigenous microorganisms and rapid extraction of microbial ADN from sediment. *Appl. Environ. Microbiol.* 60:1572-1580.

Moyer, C.L., J.M. Tiedje, F.C. Dobbs y D.M. Karl. 1998. Diversity of deep-sea hydrothermal vent *Archaea* from Loihi Seamount, Hawaii. *Deep-Sea Res. II* 45:303-317.

Muyzer, G., E.C. de-Waal y A.G. Uitterlinden. 1993. Profiling of complex microbial populations by denaturing gradient gel electrophoresis analysis of polymerase chain reaction-amplified genes coding for 16S rRNA. *Appl. Environ. Microbiol.* 59:695-700.

Nübel, U., F. García-Pichel, M. Kühl y G.Muyzer 1999a. Spatial scale and diversity of benthic cyanobacteria and diatoms in a salina. *Hydrobiologia*. 401:199-206.

Nübel, U., F. García-Pichel, M. Kühl y G. Muyzer. 1999b. Quantifying microbial diversity: morphotypes, 16S rRNA genes, and carotenoids of oxygenic phototrophs in microbial mats. *Appl Environ. Microbiol*. 65:422-430.

Øvreås L., F.L. Daae, M. Heldal, F. Rodríguez-Valera y V. Torsvik. Artículo presentado en el 9th InteRNAtional Symposium on Microbial Ecology: Interaction in the Microbial World, Amsterdam, 26 al 31 de agosto de 2001.

Øvreås L., Forney L., Daae F.D. y Torsvik V. 1997. Distribution of bacterioplancton in meromictic lake Sælevannet, as determined by denaturing gradient gel electrophoresis of PCR-amplified gene fragments coding for 16S rRNA. *Appl. Environ. Microbiol*. 63:3367-2273.

Pedersen, K. 2000. Exploration of deep intraterrestrial life. *FEMS Microbiol. Lett*. 185:9-16.

Pickup, R.W. 1991. Development of molecular methods for the detection of specific bacteria in the environment. *J.Gen. Microbiol*. 137:1009-1019.

Porteous, L.A., R.J. Seidler y L.S. Watrud. 1997. An improved method for purifying DNA from soil for polymerase chain reaction amplification and molecular ecology applications. *Mol. Ecol*. 6: 787-791.

Preston, F.W. 1948. The commonness, and rarity, of species. *Ecology*. 29:254-283.

Price, P.W. 1975. *Insect Ecology*. Wiley, New York, N.Y.

Purdy K.J., T.M. Embley, S. Takii y D.B. Nedwell. 1996. Rapid extraction of DNA and rRNA from sediments by a novel hydroxiapatite spin-column method. *Appl. Environ. Microbiol*. 62:3905-3907.

Raaijamakers, J.G.W. 1987. Statistical analysis of the Michaelis-Menten equation. *Biometrics* 43:793-803.

Ritz, K., B.S. Griffiths, V.L. Torsvik y N.B. Hendricksen. 1997. Analysis of soil and bacterioplankton community DNA by melting profiles and reassociation kinetics. FEMS *Microbiol. Lett*. 149: 151-156.

Ricklefs, R.E. y G.L. Miller 2000. *Ecology*. W.H. Freeman and Company, New York, N.Y.

Rosado, A.S., G.F. Dusrte, L. Deldin y J.D. van Elsas. 1997. Molecular microbial ecology: a mini review. *Revista de Microbiología* 28:135-147.

Rosenzweig, R.F., R.R. Sharp, D.S. Treves y J. Adams 1994. Microbial evolution in a simple unstructured environment: Genetic differentiation in *Escherichia coli*. *Genetics*. 137:903-917.

Samways, M.J. 1994. *Insect conservation biology*. Chapman & Hall, Londres.

Seber, G.A.F. 1973. *The estimation of animal abundance and related parameters*. Griffin, Londres.

Simek, K., P. Kojecka, J. Nedoma, *et al.* 1999. Shifts in bacterial community composition associated with different microzooplankton size fractions in a eutrophic reservoir. *Limnol. Oceanogr*. 44:1634-1644.

Soberón, J. y J. Llorente. 1993. The use of species accumulation functions for the

prediction of species richness. *Conserv. Biol*. 7:480-488.

Stackebrandt, E., W. Liesak y B.M. Goebel. 1993. Bacterial diversity in a soil sample of a subtropical Australian environment as determined by 16S rADN analysis. *FASEB J.* 7:232-236.

Stackebrandt, E. y F.A. Rainey. 1995. *Partial and complete 16S rADN sequences, their use in generation of 16S rADN phylogenetic trees and their implications in molecular ecological studies*. En: A.D.L. Akkermans, J.D. van Elsas y F.J. de Brujin [eds.]. Molecular Microbial Ecology Manual. Kluwer, Dordrecht.

Staley, J.T. 1997. Biodiversity: are microbial species threatened? *Curr. Opin. Biotechnol*. 8:340-345.

Steffan, R.J. y R.M. Atlas. 1988. DNA amplification to enhance detection of genetically engineered bacteria in environmental samples. *Appl. Environ. Microbiol*. 54:2185-2191.

Steffan, R.J., J. Goksoyr, A.K. Bej y R.M. Atlas. 1988. Recovery of DNA from soils and sediments. *Appl. Environ. Microbiol*. 56:776-781.

Sugihara, G. 1980. Minimal community structure: an explanation of species abundance patterns. *Am. Nat*. 116:770-787.

Suzuki, M.T. y S.J. Giovannoni. 1996. Bias caused by template annealing in the amplification mixtures of 16S rRNA genes by PCR. *Appl. Environ. Microbiol*. 62:625-630.

Tiedje, J.M. 1995. *Approaches to the comprehensive evaluation of prokaryote diversity of a habitat*. En: D. Allsopp, D.L. Hawksworth y E.R. Colwell [eds.]. CAN InteRNAtional, Microbial Diversity and Ecosystem Function. Wallingford.

Tiedje, J.M., S. Asuming-Brempong, K. Nusslein, T.L. Marsh y S.J. Flynn. 1999. Opening de black box of soil microbial diversity. *Appl. Soil Ecol*. 64:2894-2898.

Torsvik, V., J. Goksøyr y L. Daae. 1990. High diversity in DNA of soil bacteria. *Appl. Environ. Microbiol*. 56:782.

Torsvik, V., F.L. Daae y J. Goksøyr 1995. *Extraction, purification, and analysis of ADN from soil bacteria*. En: J.T. Trevors y J.D. van Elsas [eds.]. *Nucleic Acids in the environment: methods and applications*. Springer-Verlag, Berlin.

Torsvik, V., R. Sørheim y J. Goksøyr 1996. Total bacterial diversity in soil and sediment communities – a review. *J. Indust. Microbiol*. 17:170-178.

Torsvik, V., F.L. Daae, R.-A. Sandaa y L. Øvreås. 1998. Novel techniques for analysisng diversity in natural and perturbed environments. *J. Biotechnol*. 64:53-62.

Torsvik V., L. Øvreås y T.F. Thingstad. 2002. Prokaryotic diversity – Magnitude, dynamics and controlling factors. *Science* 296:1064-1066.

Tsai, Y. y B.H. Olsen. 1992. Detection of low numbers of bacterial cells in soils and sediments by polymerase chain rection. *Appl. Environ. Microbiol*. 58:754-757.

Tunlid, A., H.A.J. Hoitink, C. Low y D.C. White. 1989. Characterization of bacteria that supress Rhizotocnia damping-off in bark compost media by analysis of fatty acid biomarkers. *Appl. Environ. Microbiol*. 55:1368-1374.

Ueda, T., Y. Suga y T. Matsuguchi. 1995. Molecular phylogenetic analysis of a soil microbial community in a soybean field. *Eur. J. Soil Sci*. 46: 415-421.

van Winzingerode, F., U.B. Gobel y E. Stakebrandt. 1997. Determination of microbial diversity in environmental samples: pitfalls of PCR-based rRNA analysis. *FEMS Microbiol. Rev*. 21:213-229.

Wagner, M., A. Loy, R. Nogueira, U. Purkhold, N. Lee y H. Daims., 2002. Microbial community composition and function in waste water treatment plants. *Antoine van Leuuwenhoek*. 81: 65-680.

Wagner, R. 1994. The regulation of ribosomal RNA synthesis and bacterial cell growth. *Arch. Microbiol.* 161: 100-109.

Wallner, G., B. Fuchs, S. Spring, W.Beisker y R. Amman. 1997. Flow sorting of microorganisms for molecular analysis. *Appl. Environ. Microbiol*. 63: 4223-4231.

Ward, D.M., M.M. Bateson, R. Weller y A.L. Ruff-Roberts. 1992. *Ribosomal RNA analysis of microorganisms as they occur in nature*. En: K.C. Marshall [ed.]. *Advances in Microbial Ecology*. Plenum Press. New York.

Whitman, W.B., D.C. Coleman y W.J. Wiebe 1998. Prokaryotes: the unseen majority. *Proc. Natl. Acad. Sci. U.S.A*. 95:6578.

Widmer, F., Fliebach A., Lackzó E., Schulze-Aurich J. y Zeyer J. 2001. Assesing soil biological characteristics: a comparison of bulk soil community ADN-, PFLA-, and BIOLOG™ – analyses. *Soil Biology & Biochemestry*. 33:1029-1036.

Woese, C.R. 1987. Bacterial evolution. *Microbiol. Rev*. 51:221-271.

Woese, C.R., O. Kandler y M.L. Wheelis. 1990. Towards a natural system of organisms: proposal for the domains Archaea, Bacteria and Eucarya. *Proc. Natl. Acad. Sci. U.S.A.* 87:4576-4579.

Wullings, B.A., A.R. van Beuningen, J.D. Janse y A.D.L. Akkermans. 1998. Detection of Ralstonia solanacearum wich causes brown rot of potato, by fluorescent in situ hybridization with 23S rRNA-targeted probes. *Appl. Environ. Microbiol*. 64:4546-4554.

Young, P.1997. Major microbial diversity initiative recommended. *ASM News*.63: 417-421.

Zarda, B., D. Hahn, A. Chatzinotas, W. Schonhuber, J. Neef, R.I. Amann y J. Zeyer. 1997. Analysis of bacterial community structure in bulk soil by in situ hybridization. *Arch. Microbiol.* 168:185-192.

Zelles, L. y Q.V. Bai 1993. Fractionation of fatty acids derived from soil lipids by solid phase extraction and their quantitative analysis by GC-MS. *Soil Biology &*

Biochemestry. 25:495-507.

Zheng, D., E.W. Alm, D.A. Stahl y L. Raskin. 1996. Characterization of universal small-subunit rRNA hybridizaton probes for quantitative molecular microbial ecology studies. *Appl. Environ. Microbiol.* 62:4504-4513.

Zhou, J., M.A. Bruns y J.M. Tiedje. 1996. DNA recovery from soils of diverse composition. *Appl. Environ. Microbiol.* 62:312-322.

Zhou, J., B. Xia, D.S. Treves, L.-Y. Wu, T.L. Marsh, R.V. O'Neill, A.V. Palumbo y J.M. Tiedje. 2002. Spatial and resource factors influencing high microbial diversity in soil. *Appl. Environ. Microbiol.* 68(1):326-334.

Cuarta parte

La ecología molecular de plantas y animales

Capítulo 13

Ecología evolutiva de las zonas de hibridación

Antonio González Rodríguez

La hibridación natural puede definirse como la producción, en condiciones naturales, de descendencia a partir del entrecruzamiento de individuos pertenecientes a poblaciones que pueden distinguirse con base en por lo menos un carácter heredable, independientemente de su estatus taxonómico (Harrison, 1990, 1993; Arnold, 1997). Por extensión, un híbrido es un individuo descendiente de progenitores pertenecientes a poblaciones diferenciadas en por lo menos un carácter heredable. Es importante notar que en esta última definición quedan incluidos no únicamente los híbridos de primera generación (F_1), sino también todos los individuos con una gran variedad de genotipos recombinantes que pueden producirse en posteriores generaciones (F_2, F_3, retrocruzas, etc.). Las poblaciones naturales que, como consecuencia de este tipo de eventos, incluyen a individuos pertenecientes a ambas formas progenitoras, así como a híbridos con diversas genealogías, se conocen como *enjambres híbridos* (Arnold, 1997). Geográficamente hablando, las zonas de hibridación son las localidades específicas en las que dos o más grupos de organismos genéticamente diferenciados se encuentran y forman descendencia híbrida. Finalmente, el término introgresión se refiere al flujo de genes entre dos formas diferenciadas, como producto de repetidos eventos de hibridación y retrocruzamiento.

Los procesos de hibridación natural han llamado la atención de botánicos y zoólogos durante mucho tiempo y su estudio se ha llevado a cabo desde

varias perspectivas. Desde el punto de vista de la taxonomía y la sistemática, se han documentado miles de casos de hibridación en prácticamente todos los grupos de plantas y animales; muchos de estos híbridos han sido descritos y caracterizados morfológicamente, y, en un número menor de casos, también mediante diversos tipos de marcadores: citogenéticos, químicos y moleculares. Recientemente, la utilización de caracteres moleculares ha servido de igual forma para inferir en géneros como *Paeonia* (Sang *et al*., 1995), *Helianthus* (Rieseberg, 1991), *Gossypium* (Wendel *et al*., 1991), entre otros, la existencia de especies o grupos de especies que posiblemente se originaron a partir de un ancestro híbrido, lo cual se traduce en relaciones filogenéticas reticuladas, más que estrictamente divergentes.

Por su parte, los biólogos evolutivos han llegado a considerar a las zonas de hibridación como *laboratorios naturales* (Hewitt, 1988) o *ventanas* (Harrison, 1990) en los que es posible observar la evolución de forma privilegiada. Por ejemplo, el estudio de la hibridación es fundamental para comprender la naturaleza de las diferencias entre especies, así como los mecanismos que originan el aislamiento reproductivo. Asimismo, los notables patrones de variación geográfica que pueden encontrarse en las zonas de hibridación han despertado el interés acerca de los procesos genéticos y ecológicos que ocurren en tales zonas. En particular destacan las interrogantes en torno al origen, estructura interna, dinámica y destino de las zonas de hibridación, con muchas respuestas que no han terminado de discutirse. Por el contrario, se trata de un área, como otras en la biología evolutiva, en la que la utilización de marcadores moleculares ha resultado notablemente útil para reformular los viejos problemas, así como para acrecentar la calidad y cantidad de los datos empíricos disponibles (Rieseberg y Ellstrand, 1993; Rieseberg, 1995).

Aspectos teóricos

Los primeros modelos conceptuales sobre el origen y dinámica de las zonas de hibridación se elaboraron durante las primeras décadas del siglo XX. Dobzhansky (1940) y Mayr (1942) consideraron que las zonas de hibridación se forman entre poblaciones que se han diferenciado por aislamiento geográfico (alopatría), y que posteriormente experimentan alguna modificación en sus áreas de distribución, que ocasiona que éstas se solapen en alguna medida. Como resultado, se forman zonas que se denominan de contacto secundario, o zonas de hibridación secundarias. De acuerdo con estos autores, cuando ocurre divergencia entre poblaciones en condiciones alopátricas, se da la

adquisición, vía selección natural, de combinaciones distintas de genes que "funcionan bien juntos" y que constituyen la base de la adaptación a ambientes desiguales. Por lo tanto, la hibridación por contacto secundario no podría sino conducir a la disrupción de los genotipos coadaptados propios de los progenitores, y a la consiguiente reducción en la adecuación de los híbridos, bajo prácticamente cualquier circunstancia. Desde esta perspectiva, las zonas de hibridación son fenómenos evolutivamente transitorios, que desembocan en el reforzamiento de las barreras reproductivas (principalmente precigóticas), debido a la selección en contra de la hibridación (Howard, 1993).

Durante cierto tiempo existió consenso en que geográficamente las zonas de hibridación son fundamentalmente clinas en caracteres que distinguen a poblaciones con distribución contigua y que pueden por lo tanto caracterizarse mediante un muestreo a lo largo de un transecto perpendicular a la franja de contacto (Mayr, 1963; Endler, 1977; Barton y Hewitt, 1985; Hewitt, 1989). Sin embargo, Endler (1977) mostró que la aparición de este tipo de clinas no requiere necesariamente de la diferenciación previa de las poblaciones en aislamiento geográfico, sino que también pueden formarse zonas de hibridación primarias cuando existe un gradiente de selección capaz de producir diferenciación genética aún cuando no hay ninguna restricción al flujo génico entre poblaciones. Distinguir entre ambos escenarios históricos (*i.e.* origen primario vs. secundario) a partir del patrón contemporáneo de variación clinal en un carácter a lo largo de un solo transecto es muy difícil o imposible (Endler, 1977).

Dos líneas de evidencia se han utilizado frecuentemente como argumentos a favor de un origen por contacto secundario de una alta proporción de las zonas de hibridación. En muchas ocasiones las clinas son muy pronunciadas (el ancho de la zona de hibridación es de unas pocas decenas o cientos de metros), y se ha sugerido que la intensidad de la selección tiene que ser muy grande para producir este tipo de patrones (Endler, 1977; Barton y Hewitt, 1985). También resulta difícil usar la diferenciación *in situ* como una explicación para la variación concordante (clinas en la misma posición y con el mismo ancho) que se observa frecuentemente para caracteres de distinto tipo y en principio independientes, como pueden ser morfológicos, cromosómicos, conductuales y moleculares (Hewitt y Barton, 1980; Hewitt, 1989; Harrison, 1990). Una tercera línea de evidencia usada en algunos casos es la documentación directa de que la zona de hibridación se originó por cambios en la distribución geográfica de dos taxa después de una alteración ambiental, como las producidas por las glaciaciones (Hewitt, 1989).

Sin embargo, no todas las zonas de hibridación tienen la forma de una franja estrecha. Harrison y Rand (1989) propusieron el concepto de mosaico para aplicarlo a zonas de hibridación con una estructura geográfica más compleja en la que los recursos, hábitats o factores físicos adecuados para cada una de las formas progenitoras se encuentran entremezclados formando hábitats discontinuos. Esto ocurre, por ejemplo, en una zona de hibridación entre los grillos *Gryllus firmus* y *G. pennsylvanicus*, que se caracteriza por la estrecha asociación de cada especie con un cierto tipo de suelo. Esta observación se ha interpretado como evidencia de selección dependiente del ambiente (Harrison y Rand, 1989; Rand y Harrison, 1989). Por otro lado, caracteres morfológicos, aloenzimáticos y del ADN mitocondrial muestran idénticos patrones de variación espacial, lo que se ha usado para argumentar a favor de un origen por contacto secundario (Harrison y Rand, 1989; Rand y Harrison, 1989).

Una vez formadas, la dinámica y evolución de las zonas de hibridación dependen de varios factores. En prácticamente todos los modelos uno de los factores más importantes siempre es la naturaleza e intensidad de la selección natural sobre los híbridos. A pesar de que en varios de estos modelos se asume, como pensaron inicialmente Dobzhansky y Mayr, que los híbridos inevitablemente muestran menor adecuación que los progenitores puros está claro que el reforzamiento de las barreras reproductivas no es la única vía posible de evolución en las zonas de hibridación. De hecho, pueden existir indefinidamente aún si existe fuerte selección en contra de los híbridos cuando, por ejemplo, nuevos progenitores se dispersan constantemente hacia la zona de hibridación y se entrecruzan. Éste es en esencia el modelo de "zona de tensión" (Key, 1968; Barton y Hewitt, 1985, 1989). Su premisa fundamental es que por causas endógenas (*i.e.* disrupción de los genomas de los progenitores), la adecuación de los híbridos es reducida, por lo que la selección natural debe actuar siempre en su contra, independientemente del ambiente en el que se encuentren. Según Barton y Hewitt (1985) "la mayoría de los fenómenos conocidos como zonas de hibridación son clinas mantenidas por el balance entre la dispersión y la selección en contra de los híbridos". Este modelo predice y explica la estructura geográfica y genética que se observa en algunas zonas de hibridación bien estudiadas. El que puedan permanecer como franjas estrechas con numerosos caracteres distintos variando a lo largo de ellas según un mismo patrón difícilmente puede atribuirse a la selección producida por factores ecológicos, que en todo caso produciría clinas con distinta forma y posición para caracteres diferentes (Barton y Hewitt, 1985; Barton y Gale, 1993; Hewitt, 1988, 1989). Una de las propiedades más notables de las zonas

de tensión es que pueden moverse de lugar debido a que la selección que opera en ellas es independiente de las condiciones ambientales locales. De acuerdo con el modelo quedarán estacionadas en áreas de baja densidad poblacional, puesto que la dirección neta de dispersión de los progenitores será de las áreas de mayor densidad hacia las de menor densidad poblacional (Barton y Hewitt, 1985; Hewitt, 1988, 1989). También se puede esperar, a partir de la inmigración constante de los progenitores y de la selección en contra de los híbridos, que en las zonas de tensión exista desequilibrio de ligamiento entre los alelos provenientes de cada forma progenitora, así como desequilibrio citonuclear (Barton y Hewitt, 1985; M. Arnold, 1992; Barton y Gale, 1993; J. Arnold, 1993).

Algunos modelos alternativos han considerado otras posibles dinámicas. Existe la posibilidad de que la adecuación relativa de los híbridos sea en mayor medida un resultado de causas exógenas más que endógenas. En este caso se espera que las clinas para los caracteres bajo selección se asocien con un gradiente ambiental o con un área de transición ecológica. Otros caracteres no seleccionados seguirán patrones de variación diferentes, a menos que se encuentren ligados con los caracteres bajo selección (Barton y Hewitt, 1985; Hewitt, 1988, 1989). Si la adecuación de los híbridos es menor que la de los progenitores, este tipo de zonas de hibridación requieren, al igual que las zonas de tensión, de la inmigración constante de los progenitores para permanecer estables (Endler, 1977; Barton y Hewitt, 1985; Hewitt, 1988, 1989; Harrison, 1990). En cambio, un escenario diferente aparece cuando los progenitores muestran una menor adecuación que los híbridos en la zona de hibridación, como puede ocurrir, según algunos autores (Anderson, 1948; Heiser, 1973; Moore, 1977; Grant, 1981) en áreas ecológicamente perturbadas o ecotonos. Puesto que los híbridos son seleccionados a favor, este tipo de zonas de hibridación no dependen directamente de la dispersión de los progenitores para persistir. Sin embargo, la superioridad de los híbridos se encuentra confinada a estos hábitats particulares, por lo que no pueden establecerse fuera del área de hibridación. Por esta razón a éste se le llama el modelo de la *superioridad restringida del híbrido* (Moore, 1977).

Además de ocasionar el reforzamiento de las barreras reproductivas entre dos taxa, o de permanecer relativamente estables durante prolongados periodos de tiempo como resultado de distintas combinaciones de procesos, las zonas de hibridación pueden llevar también a la fusión de las poblaciones progenitoras o a la formación de nuevos linajes evolutivamente estables. Por ejemplo, si entre dos taxa se forma una franja angosta de hibridación por con-

tacto secundario, pero no existen diferencias en adecuación entre progenitores e híbridos, se obtendrá una clina que tenderá a difundirse gradualmente, y cuyo ancho será una función del tiempo transcurrido desde el contacto secundario y de la magnitud de la dispersión (Barton y Hewitt, 1985; Hewitt, 1988, 1989). El resultado final será una población homogénea en la que entre otras cosas, habrá aumentado la variabilidad genética.

El modelo recientemente aparecido de las novedades evolutivas (Arnold, 1997) ha hecho énfasis en las vías por medio de las cuales las zonas de hibridación pueden ser centros de origen de nuevos taxa, distintos de ambas formas progenitoras. Este tipo de modelos son muy necesarios considerando la gran cantidad de especies de animales (Bullini, 1994; Dowling y Secor, 1997) y de plantas (Rieseberg, 1997) que probablemente se originaron por procesos de hibridación. En este caso, el factor determinante es nuevamente la intensidad y naturaleza de la selección actuando sobre los híbridos. Cuando la selección endógena en contra no es demasiado fuerte (como sería el caso de completa esterilidad o inviabilidad de los híbridos), pueden ponerse de manifiesto diferencias en adecuación entre individuos híbridos. Esto, aunado a la posibilidad de formación de numerosos genotipos recombinantes en sucesivas generaciones de hibridación y retrocruza, ofrece a la selección abundante variación sobre la que actuar. De esta manera, es de esperarse que aumente la frecuencia de ciertos genotipos híbridos con adecuación particularmente alta, mismos que puedan incluso colonizar exitosamente hábitats novedosos y constituir a la larga linajes evolutivamente estables.

Estudios de zonas de hibridación en plantas y animales

Los estudios empíricos de zonas de hibridación en plantas y animales han estado en su mayor parte guiados por los modelos citados previamente. A partir de la búsqueda de patrones de variación esperados bajo diferentes escenarios, es posible realizar algunas inferencias sobre el origen y dinámica de zonas particulares. El uso de marcadores para determinar los patrones de variación genética (*i.e.* estructura genética; véase el capítulo 18 de este libro) en las zonas de hibridación, es en ocasiones la única forma posible de detectar eventos que ocurren con una frecuencia baja, o procesos que de otra forma no pueden observarse.

Una gran diversidad de zonas de hibridación ha sido analizada bajo este enfoque en numerosos grupos animales. El punto de partida de la mayor parte de los estudios es una hipótesis de hibridación entre dos taxa basada en

caracteres morfológicos. Entre los ejemplos notables se encuentran los trabajos en grillos y chapulines (Howard y Waring, 1991; Kohlmann y Shaw, 1991; Orr, 1996); hormigas (Shoemaker *et al*., 1996); chinches acuáticas (Sperling y Spence, 1991); moluscos (Bert y Arnold, 1995); peces (Forbes y Allendorf, 1991; Scribner y Avise, 1994; Childs *et al*., 1996); sapos del género *Bombina* (Szymura y Barton, 1986, 1991; MacCallum *et al*., 1998); reptiles (Sites *et al*., 1995, 1996); y varios mamíferos (Arntzen y Wallis, 1991; Bussche *et al*., 1993; Abernethy, 1994). A pesar de la disparidad de organismos estudiados, y de marcadores genéticos empleados (cromosomas, isoenzimas, microsatélites, RAPDs, RFLPs del mtADN o del nrADN y secuencias específicas), los resultados y conclusiones de estos estudios tienen puntos en común. En algunas ocasiones, los patrones de variación espacial de los marcadores han indicado con claridad zonas de hibridación con una estructura geográfica clinal (Szymura y Barton, 1986, 1991; Kohlmann y Shaw,1991; Abernethy, 1994; Orr, 1996) aunque, recientemente, el número reportado de zonas en mosaico se ha incrementado considerablemente (Howard y Waring, 1991; Sites *et al*., 1995; Shoemaker *et al*., 1996; MacCallum *et al*., 1998). Sin embargo, de forma igualmente común, se han encontrado zonas de hibridación con estructuras complejas que combinan las características de las clinas y los mosaicos o que no están bien definidas espacialmente (Arntzen y Wallis, 1991; Sperling y Spence, 1991; Bert y Arnold, 1995; Childs *et al*., 1996).

Como ya se mencionó, la dinámica y destino de las zonas de hibridación dependen en gran medida de la naturaleza e intensidad de la selección natural. Cuando predomina la selección (ya sea endógena o exógena) en contra de los híbridos y existe dispersión de progenitores puros hacia la zona de hibridación, se espera que se encuentren patrones de asociación significativa entre los alelos en distintos loci que provienen del mismo progenitor (*i.e.* desequilibrio de ligamiento nuclear y citonuclear). Por esta razón, estimar estos parámetros de la estructura genética se considera fundamental en los estudios de zonas de hibridación. En la mayor parte, aunque no en todos los estudios en los que se han buscado, se han encontrado niveles significativos de este tipo de asociaciones (Szymura y Barton, 1986; 1991; Forbes y Allendorf, 1991; Howard y Waring, 1991; Sperling y Spence, 1991; Bussche *et al*., 1993; Abernethy, 1994; Bert y Arnold, 1995; Sites *et al*., 1995; 1996; Shoemaker *et al*., 1996; MacCallum *et al*., 1998). A pesar de que esto parece indicar que los procesos arriba citados tienen preponderancia en la mayor parte de las zonas de hibridación, en algunos casos también se ha encontrado evidencia de selección a favor de los híbridos y asociaciones genotipo-ambiente (*e.g.* Bert y Arnold, 1995). En general, la impresión

que se desprende de estos trabajos, es que ninguno de los modelos teóricos por sí solo permite aún explicar completamente las situaciones reales.

Existe un número notablemente menor de trabajos sobre la estructura genética de zonas de hibridación en plantas, lo que resulta irónico, puesto que la hibridación ocurre en plantas con una frecuencia muy alta y tradicionalmente se ha considerado mucho más importante en la evolución de este grupo que en la de los animales (Dowling y Secor, 1997). La causa es que en botánica el estudio de la hibridación se ha realizado preponderantemente desde la perspectiva de la taxonomía y la sistemática. Sin embargo, recientemente, se han puesto a prueba las predicciones de los modelos en zonas de hibridación entre especies de los géneros *Aesculus* (dePamphilis y Wyatt, 1990); *Iris* (Cruzan y Arnold, 1993, 1994; Young, 1996; Burke *et al*., 2000); *Popolus* (Paige *et al*., 1991); *Artemisia* (Freeman *et al*., 1995; Graham *et al*., 1995; Wang *et al*., 1997); *Salix* (Hardig *et al*., 2000); y *Helianthus* (Beckstrom-Sternberg *et al*., 1991; Dorado *et al*., 1992; Rieseberg *et al*., 1998). De acuerdo con algunos autores, las zonas de hibridación en plantas se caracterizan generalmente, en comparación con los animales, por la frecuente formación de enjambres híbridos y por una estructura geográfica menos definida (Paige *et al*., 1991; Harrison, 1993). No obstante, en ocasiones se ha reportado variación clinal, a veces muy marcada y a escalas espaciales que van desde unos cuantos metros hasta los cientos de kilómetros (Heywood, 1986; dePamphilis y Wyatt, 1990; Cruzan y Arnold, 1993; Graham *et al*., 1995; Young, 1996; Villani *et al*., 1999). La estructura genética se caracteriza por desequilibrio de ligamiento nuclear y citonuclear (Paige *et al*., 1991; Cruzan y Arnold, 1993; 1994, Rieseberg *et al*., 1996; Villani *et al*., 1999). Las asociaciones genotipo-ambiente, descubiertas en ocasiones, ponen de manifiesto la clara influencia de la selección exógena sobre la estructura de algunas zonas (Heywood, 1986; Cruzan y Arnold, 1993).

Las plantas poseen por lo general dos vías por las que puede ocurrir flujo génico: la dispersión de polen y la dispersión de semillas. Los genomas citoplásmicos (*e.g.* mitocondrias y cloroplastos) se heredan a través del progenitor materno en la mayor parte de las angiospermas, lo que significa que su dispersión puede ocurrir únicamente a través de las semillas (véase el capítulo 6 de este libro). Por esta razón, el uso simultáneo de marcadores nucleares y citoplásmicos permite detectar posibles asimetrías en la dirección en que ocurre hibridación (*e.g.* cuando una de las "formas" actúa con mayor frecuencia como progenitor materno o paterno que la otra). De la misma manera, cuando existe introgresión se puede evaluar la medida relativa en que la dispersión de semillas y de polen contribuye al flujo génico entre ambos taxa.

Los estudios filogeográficos en plantas, recientemente con gran auge, han mostrado que el uso de marcadores citoplásmicos (principalmente del ADN de cloroplasto, cpADN), además de permitir inferencias sobre patrones históricos de migración, colonización y demografía (McCauley, 1995; Ennos *et al.*, 1999) pueden también revelar eventos históricos y contemporáneos de hibridación e introgresión difíciles de detectar mediante análisis morfológicos o marcadores nucleares (Rieseberg y Soltis, 1991; Rieseberg *et al.*, 1996). En grupos como *Eucalyptus* (McKinnon *et al.*, 2001, 2004), *Packera* (Bain y Jansen, 1996; Golden y Bain, 2000) y *Lithocarpus* (Cannon y Manos, 2003), se ha encontrado que las especies comparten los mismos haplotipos de cpADN con una alta frecuencia. En tales grupos, las barreras reproductivas interespecíficas se encuentran pobremente desarrolladas, y la escasa diferenciación en el cpADN se ha atribuido comúnmente a la hibridación y posterior introgresión citoplásmica, aunque en algunos casos puede tratarse también de polimorfismos ancestrales compartidos (Golden y Bain, 2000; Byrne *et al.*, 2002). En ocasiones se ha documentado el caso extremo de la completa sustitución de los genomas citoplásmicos de un taxón por los genomas citoplásmicos de un segundo taxón con el cual hibrida, fenómeno que se ha llamado "captura citoplásmica" (Rieseberg y Soltis, 1991; Rieseberg *et al.*, 1996). Por último, la distribución de los haplotipos citoplásmicos dentro y fuera de una zona de hibridación permite en ocasiones, junto con otros datos, inferir su probable origen. Esto se debe a que en una zona secundaria se puede esperar la presencia de los haplotipos citoplásmicos propios de ambas o de una de las formas progenitoras, mientras que en una zona primaria se espera la presencia de haplotipos ancestrales a los haplotipos que se encuentran fuera de la zona de hibridación (Beckstrom-Strenberg *et al.*, 1991). Este enfoque permitió recientemente establecer que una clina morfológica formada supuestamente por el contacto secundario entre *Eucalyptus populnea* y *E. brownii* es en realidad una zona de intergradación primaria (Holman *et al.*, 2003).

Bibliografía

Abernethy, K. 1994. The establishment of a hybrid zone between red and sika deer (genus *Cervus*). *Molecular Ecology* 3:551-562.

Anderson, E. 1948. Hybridization of the habitat. *Evolution* 2:1-9.

Arnold, J. 1993. Cytonuclear disequilibria in hybrid zones. *Annual Review of Ecology and Systematics* 24:521-554.

Arnold, M.L. 1992. Natural hybridization as an evolutionary process. *Annual Review of Ecology and Systematics* 23: 237-261.

Arnold, M.L. 1997. *Natural Hibridization and Evolution*. Oxford University Press, Nueva York.

Arntzen, J.W. y G.P. Wallis. 1991. Restricted gene flow in a moving hybrid zone of the newts *Triturus cristatus* and *T. marmoratus* in western France. *Evolution* 45:805-826.

Bain, J.F. y R.K. Jansen. 1996. Numerous chloroplast DNA polymorphisms are shared among different populations and species in the aureoid *Senecio* (*Packera*) complex. *Canadian Journal of Botany* 74:1719-1728.

Barton, N.H. y K.S. Gale. 1993. Genetic analysis of hybrid zones. En: R.G. Harrison [ed.]. *Hybrid zones and the Evolutionary Process*. Oxford University Press.

Barton, N.H. y G.M. Hewitt. 1985. Analysis of hybrid zones. *Annual Review of Ecology and Systematics* 16:113-148.

Barton, N.H. y G.M. Hewitt. 1989. Adaptation, speciation and hybrid zones. *Nature* 341:497-503.

Beckstrom-Sternberg, S.M., L.H. Rieseberg y K. Doan. 1991. Gene lineage analysis in populations of *Helianthus niveus* and *H. petiolaris* (Asteraceae). *Plant Systematics and Evolution* 175:125-138.

Bert, T.M. y W.S. Arnold. 1995. An empirical test of predictions of two competing models for the maintenance and fate of hybrid zones: both models are supported in a hard-clam hybrid zone. *Evolution* 49:276-289.

Bullini, L. 1994. Origin and evolution of animal hybrid species. *Trends in Ecology and Evolution* 7:422-426.

Burke, J.M., M.R. Bulger, R. A. Wesselingh y M.L. Arnold. 2000. Frequency and spatial pattering of clonal reproduction in Louisiana iris hybrid populations. *Evolution* 54:137-144.

Bussche, R.A., R.K. Chesser, M.J. Hamilton, R.D. Bradley, C.A. Porter y R.J. Baker. 1993. Maintenance of a narrow hybrid zone in *Peromyscus leucopus*: a test of alternative models. *Journal of Mammalogy* 74:832-845.

Byrne, M., B. Macdonald y D. Coates. 2002. Phylogeographical patterns in chloroplast DNA variation within the *Acacia acuminata* (Leguminosae: Mimosoideae) complex in Western Australia. *Journal of Evolutionary Biology* 15:576-587.

Cannon, C.H. y P.S. Manos. 2003. Phylogeography of the Southeast Asian stone oaks (*Lithocarpus*). *Journal of Biogeography* 30:211-226.

Childs, M.R., A.E. Echelle y T.E. Dowling. 1996. Development of the hybrid swarm between pecos pupfish (Cyprinodontidae: *Cyprinodon pecosensis*) and sheepshead

minnow (*Cyprinodon variegatus*): a perspective from allozymes and mtDNA. *Evolution* 50:2014-2022.

Cruzan, M.B. y M.L. Arnold. 1993. Ecological and genetic associations in an *Iris* hybrid zone. *Evolution* 47:1432-1445.

Cruzan, M.B. y M.L. Arnold. 1994. Assortative mating and natural selection in an *Iris* hybrid zone. *Evolution* 48:1946-1958.

dePamphilis, C.W. y R. Wyatt. 1990. Electrophoretic confirmation of interespecific hybridization in *Aesculus* (Hippocastanaceae) and the genetic structure of a broad hybrid zone. *Evolution* 44:1295-1317.

Dobzhansky, T. 1940. Speciation as a stage in evolutionary divergence. *American Naturalist* 74:312-321.

Dorado, O., L.H. Rieseberg y D.M. Arias. 1992. Chloroplast DNA introgression in southern California sunflowers. *Evolution* 46:566-572.

Dowling, T.E. y C.L. Secor. 1997. The role of hybridization and introgression in the diversification of animals. *Annual Review of Ecology and Systematics* 28:593-619.

Endler, J.A. 1977. *Geographic variation, speciation and clines*. Princeton University Press, Nueva Jersey.

Ennos, R.A., W.T. Sinclair, X.-S. Hu y A. Langdon. 1999. Using organelle markers to elucidate the history, ecology and evolution of plant populations. En: P.M. Hollingsworth, R.M. Bateman y R.J. Gornall [eds.]. *Molecular Systematics and Plant Evolution*. Taylor & Francis Ltd., Londres.

Forbes, S.H. y F.W. Allendorf. 1991. Associations between mitochondrial and nuclear genotypes in cutthroat trout hybrid swarms. *Evolution* 45:1332-1349.

Freeman, D.C., J.H. Graham, D.W. Byrd, E.D. McArthur y W.A. Turner. 1995. Narrow hybrid zone between two subspecies of big sagebrush *Artemisa tridentata* (Asteraceae). III. Developmental instability. *American Journal of Botany* 82:114-1152.

Golden, J.L. y J.F. Bain. 2000. Phylogeographic patterns and high levels of chloroplast DNA diversity in four Packera (Asteraceae) species in southwestern Alberta. *Evolution* 54: 1566–1579.

Graham, J.H., D.C. Freeman y E.D. McArthur. 1995. Narrow hybrid zone between two subspecies of big sagebrush (*Artemisia tridentata*: Asteraceae). II. Selection gradients and hybrid fitness. *Americam Journal of Botany* 82:709-716.

Grant, V. 1981. *Plant speciation*. Columbia University Press, Nueva York.

Hardig, T.M., S.J. Brunsfeld, R.S. Fritz, M. Morgan y M. Orians. 2000. Morphological evidence for hybridization and introgression in a willow (*Salix*) hybrid zone. *Molecular Ecology* 9:9-24.

Harrison, R.G. 1990. Hybrid zones: windows on evolutionary process. En: D. Futuyma y J. Antonovics [eds.]. *Oxford Surveys in Evolutionary Biology* 7:69-128.

Harrison, R.G. 1993. Hybrids and hybrid zones: Historical Perspective. *En*: R.G. Harrison, (ed.). *Hybrid zones and the Evolutionary Process*. Oxford University Press.

Harrison, R.G. y D.M. Rand. 1989. Mosaic hybrid zones and the nature of species-boundaries. En: D. Otte y J. Endler [eds.]. *Speciation and its consequences*. Sinauer Associates, Sunderland, Massachusetts.

Heiser, C.B. 1973. Introgression re-examined. *The Botanical Review* 39:347-366.

Hewitt, G.M. 1988. Hybrid zones –natural laboratories for evolutionary studies. *Trends in Ecology and Evolution* 3:158-167.

Hewitt, G.M. 1989. The subdivision of species by hybrid zones. En: D. Otte y J. Endler [eds.]. *Speciation and its consequences*. Sinauer Associates, Sunderland, Massachusetts.

Hewitt, G.M. y N.H. Barton. 1980. The structure and maintenance of hybrid zones as exemplified by *Podisma pedestris*. En: R. Blackman, G.M. Hewitt, y M. Ashburner [eds.]. *Insect cytogenetics*. Royal Entomological Society of London Symposia 10. Blackwell, Oxford.

Heywood, J.S. 1986. Clinal variation associated with edaphic ecotones in hybrid populations of *Gaillardia pulchella*. *Evolution* 40:1132-1140.

Holman, J.E., J.M. Hughes y R.J. Fensham. 2003. A morphological cline in *Eucalyptus*: a genetic perspective. *Molecular Ecology* 12:3013-3025.

Howard, D.J. 1993. Reinforcement: origin, dynamics and fate of an evolutionary hypothesis. En: R.G. Harrison [ed.]. *Hybrid zones and the evolutionary process*. Oxford University Press.

Howard, D.J. y G.L. Waring. 1991. Topographic diversity, zone width, and the strength of reproductive isolation in a zone of overlap and hybridization. *Evolution* 45:1120-1135.

Key, K.H.L. 1968. The concept of stasipatric speciation. *Systematic Zoology* 17:14-22.

Kohlmann, B. y D. Shaw. 1991. The effect of a partial barrier on the movement of a hybrid zone. *Evolution* 45:1606-1617.

MacCallum, C.J., B. Nürnberger, N.H. Barton y J.M. Szymura. 1998. Habitat preference in the *Bombina* hybrid zone in Croatia. *Evolution* 52:27-239.

Mayr, E. 1942. *Systematics and the Origin of Species.* Columbia University Press, NuevaYork.

Mayr, E. 1963. *Animal Species and Evolution*. Belknap Press, Cambridge, Massachussets.

McCauley, D.E. 1995. The use of chloroplast DNA polymorphism in studies of gene flow in plants. *Trends in Ecology and Evolution* 10:198-202.

McKinnon, G.E., R.E. Vaillancourt, H.D. Jackson y B.M. Potts. 2001. Chloroplast sharing in the Tasmanian eucalypts. *Evolution* 55:703–711.

McKinnon, G.E., G.J. Jordan, R.E. Vaillancourt, D.A. Steane y B.M. Potts. 2004. Glacial refugia and reticulate evolution: the case of the Tasmanian eucalypts. *Philosophical Transactions of the Royal Society of London series B* 359:275–284.

Moore, W.S. 1977. An evaluation of narrow hybrid zones in vertebrates. *Quaterly Review of Biology* 52:263-277.

Orr, M.R. 1996. Life-history adaptation and reproductive isolation in a grasshopper hybrid zone. *Evolution* 50:704-716.

Paige, K.N., W.C. Capman y P. Jenneten. 1991. Mitochondrial inheritance across a cottonwood hybrid zone: citonuclear disequilibria and hybrid zone dynamics. *Evolution* 45:1360-1369.

Rand, D.M. y R.G. Harrison. 1989. Ecological genetics of a mosaic hybrid zone: mitochondrial, nuclear, and reproductive differentiation of crickets by soil type. *Evolution* 43:432-449.

Rieseberg, L.H. 1991. Homoploid reticulate evolution in *Helianthus* (Asteraceae): evidence from ribosomal genes. *American Journal of Botany* 78:1218-1237.

Rieseberg, L.H. 1995. The role of hybridization in evolution: old wine in new skins. *American Journal of Botany* 82: 944-953.

Rieseberg, L.H. 1997. Hybrid origins of plant species. *Annual Review of Ecology and Systematics* 28:359-389.

Rieseberg, L.H., J. Whitton y C.R. Linder. 1996. Molecular marker incongruence in plant hybrid zones and phylogenetic trees. *Acta Botanica of the Netherlands* 45:243-262.

Rieseberg, L.H., S. J.E. Baird y A.M. Desrochers. 1998. Patterns of mating in wild sunflowers hybrid zones. *Evolution* 52:713-726.

Rieseberg, L.H. y D.E. Soltis. 1991. Phylogenetic consequences of cytoplasmic gene flow in plants. *Evolutionary Trends in Plants* 5:65-84.

Rieseberg, L.H. y N.C. Ellstrand. 1993. What can molecular and morphological markers tell us about plant hybridization? *Critical Reviews in Plant Sciences* 12:213-241.

Sang, T., D.J. Crawford y T.F. Stuessy. 1995. Documentation of reticulate evolution in peonies (*Paeonia*) using internal transcribed spacer sequences of nuclear ribosomal DNA: Implications for biogeography and concerted evolution. *Proc. Natl. Acad. Sci. USA* 92:6813-6817.

Scribner, K.T. y J.C. Avise. 1994. Population cage experiments with a vertebrate: the temporal demography and citonuclear genetics of hybridization in *Gambusia* fishes. *Evolution* 48:155-171.

Shoemaker, D., K.G. Ross y M.L. Arnold. 1996. Genetic structure and evolution of a fire ant hybrid zone. *Evolution* 50:1958-1976.

Sites, J.W., C.J. Basten y M.A. Asmussen. 1996. Cytonuclear genetic structure of a hybrid zone in lizards of the *Sceloporus grammicus* complex (Sauria, Phrynosomatidae). *Molecular Ecology* 5:379-392.

Sites, J.W., N.H. Barton y K.M. Reed. 1995. The genetic structure of a hybrid zone between two chromosome races of the *Sceloporus grammicus* complex (Sauria, Phrynosomatidae) in central Mexico. *Evolution* 49:9-36.

Sperling, F.H. y J.R. Spence. 1991. Structure of an asymmetric hybrid zone between two water strider species (Hemiptera: Gerridae: *Limnoporus*). *Evolution* 45:1370-1383.

Szymura, J.M. y N.H. Barton. 1986. Genetic analysis of a hybrid zone between the fire-bellied toads, *Bombina bombina* and *Bombina variegata*, near Cracow in southern Poland. *Evolution* 40:1141-1159.

Szymura, J.M. y N.H. Barton. 1991. The genetic structure of the hybrid zone between the fire-bellied toads *Bombina bombina* and *B. variegata*: comparisons between transects and between loci. *Evolution* 45:237-261.

Villani, F., A. Sansotta, M. Cherubini, D. Cesaron y V. Sbordoni. 1999. Genetic structure of natural populations of *Castanea sativa*: evidence of a hybrid zone. *Journal of Evolutionary Biology* 12:233-244.

Wang, H., E.D. McArthur, S.C. Sanderson, J.H. Graham y D.C. Freeman. 1997. Narrow hybrid zone between two subspecies of big sagebrush (*Artemisia tridentata*: Asteraceae). IV. Reciprocal transplant experiments. *Evolution* 51:95-102.

Wendel, J.F., Stewart, J. McD. y J.H. Rettig. 1991. Molecular evidence for homoploid reticulate evolution among Australian species of *Gossuypium*. *Evolution* 45:694-711.

Young, N.D. 1996. Concordance and discordance: A tale of two hybrid zones in the Pacific Coast Irises (Iridaceae). *American Journal of Botany* 83:1623-1629.

Capítulo 14

Filogeografía y vertebrados

Ella Vázquez Domínguez

Many species have proved to exhibit a deep and geographically structured mtDNA phylogenetic history. Study of the relationship between genealogy and geography constitutes a discipline that can be termed intraspecific phylogeography.

Avise *et al.*, 1987

En la historia de la ciencia existen palabras o frases sobre un concepto que se convierten en parte del léxico de un área particular de estudio, por ejemplo biodiversidad, biogeografía de islas o macroecología. Inician como una palabra utilitaria que ayuda a expresar ideas y, más tarde, se convierten en definiciones específicas que estimulan campos nuevos de estudio. La filogeografía pasó precisamente por un proceso similar. John Avise, de la Universidad de Georgia, y sus colaboradores, introdujeron la palabra formalmente en 1987, momento en el que resumía una observación evidente y directa de sus estudios de más de una década: las ramas de algunos árboles de genes intraespecíficos mostraban un patrón geográfico impresionante, por lo que estas relaciones entre genealogías de genes y geografía podían referirse simplemente como patrones filogeográficos (Avise *et al.*, 1987).

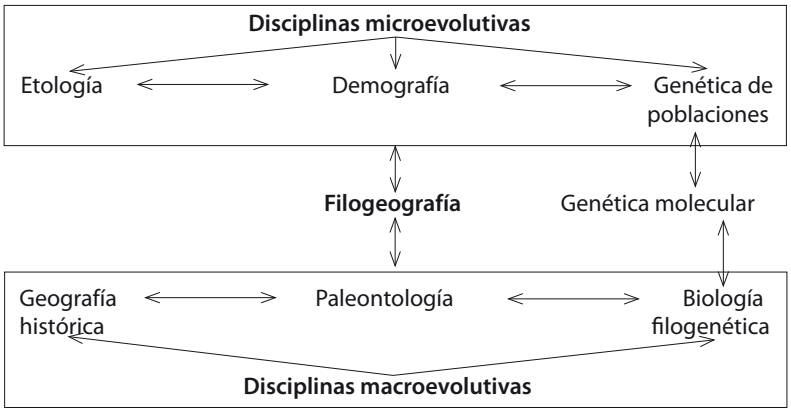
Algunos aspectos generales sobre filogeografía

La filogeografía, como la entendemos hoy, se enmarca en un área más amplia denominada ecología molecular, que se define de manera general como el conocimiento y aplicación de marcadores genéticos moleculares para explorar preguntas y problemas en ecología y evolución. La naturaleza y la escala de los estudios de ecología molecular están definidas por el hecho de que se estudian las relaciones genéticas entre individuos, poblaciones y especies y que se emplean marcadores genéticos moleculares, ya sea en forma de proteínas o de secuencias de ADN. Dado que dichos marcadores se asocian a las genealogías, la ecología molecular también considera principios de genética y de herencia. Asimismo, en esta disciplina es fundamental contar con información cualitativa y cuantitativa sobre los cambios en la estructura o composición genética, ya sea entre generaciones a lo largo del tiempo, o espacialmente entre individuos, poblaciones o especies (Carvalho, 1998; Vázquez-Domínguez, en prensa). La filogeografía trabaja con los componentes históricos o filogenéticos de la distribución espacial de linajes de genes, y considera como ejes el tiempo y el espacio, en los cuales (idealmente) se mapean las genealogías de estudio, por lo que la definición que más me gusta es la que su creador menciona "... es el campo de estudio relacionado con los principios y procesos que gobiernan la distribución geográfica de linajes de genes, sobre todo aquellos entre y dentro de especies cercanamente relacionadas" (Avise, 2000).

Se ha situado a la filogeografía como punto de unión o camino medio crítico entre la macroevolución y la microevolución. En el momento en que surge la filogeografía ya existían campos bien desarrollados sobre el estudio micro y macroevolutivo. Sin embargo, continuaba el debate de qué tanto era posible extrapolar los procesos microevolutivos que operan dentro de las especies para explicar las diferencias macroevolutivas entre especies (Gould, 1980; Mayr, 1982). La filogeografía permite y refuerza la interacción entre estos dos niveles, donde la genética molecular juega un papel primordial como herramienta indispensable (figura 1) (Avise *et al.*, 1987; Vázquez-Domínguez, en prensa).

La filogeografía enfatiza aspectos históricos de la distribución espacial actual de linajes de genes utilizando información temporal (histórica) que ha sido posible obtener a partir del uso y análisis de la información contenida en las moléculas de ADN. Esto es, trata de interpretar la extensión y el modo en que los procesos demográficos históricos han dejado marcas evolutivas en la distribución geográfica actual de caracteres genéticamente basados

Figura 1. La filogeografía como punto de unión entre disciplinas micro y macroevolutivas, así como las principales áreas que la componen y definen (modificado de Avise, 2000)



(Avise, 2000). En otra escala (ecológica), la ecogeografía enfatiza el papel de las presiones o factores ecológicos contemporáneos en la conformación de la distribución espacial de caracteres. Así, la filogeografía permite también ubicar perspectivas tradicionales ecogeográficas en un contexto temporal más amplio, contribuyendo a expandir y balancear las perspectivas de la ecogeografía sin ser, por supuesto, excluyentes. Sirve también como balance conceptual o conciliador entre los tradicionales escenarios históricos alternativos para explicar el arreglo espacial de los organismos y sus características, básicamente entre las propuestas de vicarianza y dispersión. Estas ideas se han presentado por mucho tiempo como explicaciones opuestas del origen de taxones espacialmente separados, pero su papel relativo puede ser ponderado bajo la óptica filogeográfica; con ello ha sido posible demostrar que la distribución y estructuración de las poblaciones puede responder tanto a procesos vicariantes como de dispersión, operando en tiempos, sitios y escalas diferentes.

Y por supuesto, para el análisis e interpretación de las distribuciones de linajes se requiere de la contribución teórica, metodológica y/o conceptual de muy diversos campos, como la genética de poblaciones, genética molecular, demografía, etología, filogenia, paleontología, geología y geografía histórica.

La molécula estrella, el ADN mitocondrial

Parte de la historia del surgimiento de la filogeografía viene entrelazada no sólo con el desarrollo de técnicas moleculares, sino del conocimiento de la molécula de ADN mitocondrial (mtADN) a mediados de la década de los 70. En su trabajo de 1987, Avise y colaboradores describen las características que que idealmente debería tener una molécula para análisis filogenéticos, y reconocen que las propiedades moleculares del ADN mitocondrial de los animales superiores se acerca bastante a dicha molécula (Avise *et al.*, 1987).

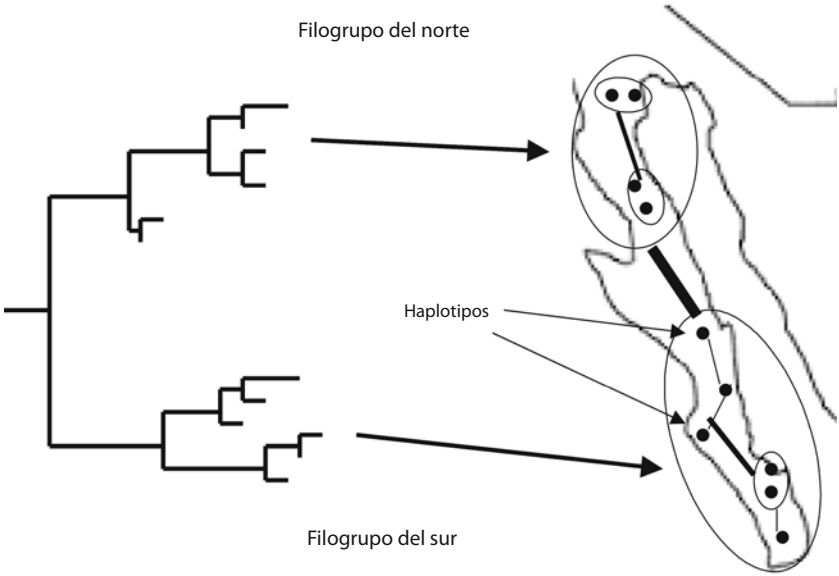
Entre las características moleculares que hacen especial al mtADN están que es una molécula circular covalentemente cerrada, de tamaño pequeño (*ca.* 16-20 kilobases), conformada por un total de 37 genes (13 ARN mensajeros, 2 ARN ribosomales y 22 ARN de transferencia), además de una región conocida, en vertebrados y equinodermos, como región control (*ca.* 1 kb) o *d-loop* (fragmento más pequeño dentro de la región control misma), que controla la replicación y transcripción en la molécula y tiene una tasa excepcionalmente alta de sustitución y de polimorfismo en muchos taxones (para una descripción detallada del mtADN ver Brown, 1985; Wainscoat, 1987; Lunt *et al.*, 1998). Las propiedades más interesantes en términos filogeográficos son su alta tasa de evolución (sustitución) a nivel de secuencias de nucleótidos, su prácticamente nula recombinación, gran variación intraespecífica, y más importante, su herencia estrictamente materna (con muy escasas excepciones).

En filogeografía se aplican los mismos algoritmos que se emplean para reconstruir árboles de especies o taxones superiores, dando como resultado un cladograma, llamado filograma, que representa las relaciones genealógicas entre haplotipos y donde las ramas de diferentes longitudes expresan la cantidad de cambio evolutivo (en términos, por ejemplo, de número de pasos mutacionales o número de sustituciones; figura 2 (véase el capítulo 4 de este libro).

Las variantes o haplotipos del mtADN registran la historia matrilineal de eventos mutacionales, y pueden conectarse filogenéticamente en un filograma, el cual puede a su vez sobreponerse a la distribución geográfica del grupo de estudio, a fin de interpretar los procesos evolutivos responsables de la distribución actual de la especie o grupo de especies (Avise, 2000).

Dados los cientos o miles de moléculas de mtADN que tiene un individuo, la heteroplasmia (presencia de dos o más haplotipos mitocondriales dentro de una misma célula) parecía ser un problema potencial en estudios con mtADN. Contrario a lo esperado, se conoce que los casos de hetero-

Figura 2. Filogeografía hipotética de una especie cuya distribución abarca la península de Baja California. En el filograma, son evidentes 2 filogrupos altamente divergentes, que tienen concordancia con la distribución geográfica de las poblaciones en el norte y el sur de la península. Los puntos indican haplotipos diferentes, y el grosor de la línea que los une el número de pasos mutacionales (nótese el máximo grosor en la línea que une haplotipos del norte y del sur)



plasmia son mínimos y no representan un problema en estudios rutinarios con mtADN en animales. Por el contrario, se reconoce que existe la posibilidad de efectos de cambios homoplásicos (caracteres similares pero no derivados de un ancestro común) en estudios con mtADN (*e.g.* cambios paralelos o convergentes en sitios de restricción), sobre todo en términos de ambigüedad en la construcción de los árboles o redes de haplotipos. Sin embargo, existen formas de evaluar y lidiar con la homoplasia (ver Templeton, 1983; véase también el capítulo 4 de este libro).

Al comparar secuencias homólogas, el número de diferencias en las bases es un estimador del tiempo de divergencia entre linajes, por lo que la filogeografía emplea estimaciones del reloj molecular, ya sea de mtADN o de marcadores nucleares. El mtADN tiene una tasa de diferenciación alta (estimada en aproxi-

madamente 2% de divergencia de secuencias por cada millón de años (ma) en mamíferos; por ejemplo Brown *et al.*, 1979), pero atribuible sólo a cambios en ciertas posiciones de nucleótidos específicas. Asimismo, diferentes regiones y genes presentan tasas de evolución variadas dentro de un linaje. Por ejemplo, la región control ha sido particularmente útil para análisis filogeográficos en escalas muy pequeñas de tiempo microevolutivo (*e.g.*, miles o cientos de miles de años), ya que tiene una tasa excepcionalmente alta de sustitución de nucleótidos y altos niveles de polimorfismo intraespecífico (Lunt *et al.*, 1998). Así, los estudios de comparaciones filogenéticas con mtADN, deben en general concretarse a poblaciones coespecíficas y de especies cercanamente relacionadas, cuyas separaciones no vayan más allá de algunos millones de años (convencionalmente alrededor de 5 ma).

La filogenia de genes y la demografía de poblaciones están íntimamente relacionadas, de manera que las variables demográficas son factores decisivos que gobiernan la profundidad, la forma y los patrones demográficos de los árboles de genes (para una revisión detallada sobre la relación demografía-filogeografía ver Avise, 2000). Asimismo, sabiendo que hacia atrás en el tiempo, los haplotipos de genes mitocondriales o nucleares coalescen en un ancestro común, es fácil entender la homología entre las rutas o árboles coalescentes del mtADN de herencia uniparental (materna) y los árboles genealógicos familiares donde vemos el camino que han seguido los apellidos de nuestros antepasados (en este caso generalmente es la línea paterna). Es la teoría de coalescencia la que describe las relaciones históricas entre linajes de genes como una función de los parámetros demográficos; es también el nombre que actualmente se le da al análisis matemático y estadístico formal de las genealogías de genes (Avise, 2000; Cruzan y Templeton, 2000; Templeton, 2004; véase el capítulo 4 de este libro).

Es importante mencionar que los estudios filogeográficos se han hecho casi en su totalidad con ADN citoplásmico (mitocondrial y de cloroplasto), aunque el uso de haplotipos nucleares es cada vez más frecuente y exitoso conforme se han desarrollado las técnicas moleculares y de interpretación. Ello ha sido motivado también como resultado de querer evaluar modelos históricos más sofisticados, lo que exige la utilización de diferentes marcadores complementarios que permitan descifrar la historia completa de los patrones genealógicos observados en las poblaciones naturales (Carbone y Kohn, 2001; Hare, 2001). El uso de loci nucleares diploides en análisis filogeográficos conlleva complicaciones, como la influencia de la selección y el factor temporal en los patrones estudiados, pero varios estudios han demostrado la factibilidad de análisis e interpretación filogeográfica (ver referencias en Hare, 2001).

Así, para concluir esta breve introducción a la teoría, podemos mencionar algunos aspectos fundamentales del estudio filogeográfico: la unidad de estudio es el individuo, y más estrictamente, el haplotipo; se evalúan polimorfismos que son únicos –que no se repiten– y definen claramente la filogenia de los mismos; utiliza árboles de genes, es decir, con base en haplotipos, y no árboles de especies (o taxones o áreas), con los que puede trazarse la filogenia matrilineal en el caso del ADN mitocondrial (haploide), así como las genealogías de genes nucleares (diploides); y se utilizan enfoques basados en los principios de coalescencia para evaluar formal y estadísticamente las predicciones filogeográficas.

Teoría de coalescencia y métodos de análisis

Bajo neutralidad, en una población y a lo largo de las generaciones, surgen de forma permanente nuevos alelos por mutación y se pierden otros por deriva génica, de tal forma que todos los alelos de un gen en una generación derivan de –o coalescen hacia– un único alelo ancestral. Los procesos de coalescencia en poblaciones naturales "garantizan" las conexiones filogenéticas entre genotipos dentro de una especie, a través de sendas verticales de descendencia. De esta forma, al ver hacia atrás en el tiempo, los haplotipos de mtADN o de cualquier gen nuclear, eventualmente coalescen en el ancestro común. El tiempo de coalescencia depende de procesos poblacionales, como fluctuación del tamaño poblacional, selección natural o variaciones en el flujo génico (Harding, 1996). Para el caso que nos ocupa –la filogeografía– la teoría de colascencia moderna ha permitido evaluar los procesos de ramificación de árboles de genes, considerando explícitamente la historia genealógica de los alelos (Hudson, 1990, 1998; Nordborg, 2000).

Los métodos moleculares actuales permiten no sólo la estimación de las frecuencias alélicas, sino también de las relaciones genealógicas de los alelos, lo que permite considerar el flujo génico histórico y la fragmentación de poblaciones en un marco conceptual más histórico y geográfico. Sin embargo, dado que los métodos tradicionales de la genética de poblaciones no consideran los factores históricos, éstos deben adecuarse para poder hacer uso de la información histórica contenida en los estimadores moleculares. Wright (1931, 1943) introdujo los estadísticos *F*, que han permitido cuantificar la subdivisión –estructura– de las poblaciones, así como el flujo génico, a partir de las frecuencias alélicas obtenidas en diferentes localidades. Aunque estos estadísticos no hacen uso de la información temporal, nuevos procedimientos estadísticos basados en la información de ADN, pueden separar los eventos

históricos de los asociados a la estructura poblacional, por lo que actualmente se tiene no sólo la escala geográfica, sino una dimensión más: el tiempo evolutivo (Hudson *et al.*, 1992; Templeton *et al.*, 1995; Templeton, 2004).

La simple sobreposición de los árboles de haplotipos sobre la distribución geográfica de las poblaciones es una forma fácil y directa de aproximarse a los patrones filogeográficos que presenta nuestro sistema de estudio (ello puede corroborarse al observar muchas de las figuras que se presentan en publicaciones de estudios filogeográficos), y puede sugerirnos los eventos o procesos filogeográficos que se sucedieron o por el contrario dar pie a interpretaciones erradas; sin embargo, ello no constituye una validación de las hipótesis filogeográficas posibles.

Recientemente se han desarrollado métodos para examinar formalmente el efecto de los patrones históricos de migración y dispersión sobre la distribución de genes, integrando la información genealógica y de distribución para hacer inferencias sobre los patrones históricos de flujo génico (*e.g.* Hudson, 1990; Templeton *et al.*, 1995; Avise, 2000; Cruzan y Templeton, 2000; Templeton, 2004).

Los modelos recientes de genética de poblaciones que basados en la teoría de coalescencia proveen un marco conceptual estadístico para el análisis de información filogeográfica (*e.g.* Modelo de Niegel; Niegel y Avise, 1993; Avise, 2000), se ha compilado bajo el concepto de *estadística filogeográfica* (Knowles y Maddison, 2002). Una metodología que se utiliza con considerable frecuencia en filogeografía es el análisis de clados anidados (NCA, por sus siglas en inglés; Templeton *et al.*, 1995; véase también el capítulo 2 de este libro). Su popularidad se debe a tres razones principales: a) permite hacer un análisis estadístico objetivo con el que se puede evaluar la hipótesis nula de que los haplotipos de los clados se distribuyen geográficamente al azar (*i.e.* no existe asociación entre la variación haplotípica y la geografía); b) permite detectar y probar estadísticamente los mecanismos evolutivos responsables de la distribución espacial de los patrones de variación genética observada; y c) interpreta los patrones significativos utilizando criterios explícitos, con los que es posible evaluar si el muestreo ha sido adecuado para detectar una asociación significativa entre los clados y la geografía, así como para interpretar biológicamente dicha asociación. Finalmente, aquellas asociaciones significativas entre haplotipos y su distribución geográfica se pueden explicar con base en tres procesos: a) aislamiento por distancia, debido a flujo génico restringido; b) fragmentación histórica; y c) expansión del área de distribución de las poblaciones, incluyendo eventos de colonización a distancia. La descripción detallada del análisis de clados anidados, así como las herramientas

para llevarlo al cabo, se describen en Templeton *et al.* (1995) y Templeton (1998, 2004). El NCA tiene ciertas limitaciones, como que es sensible a la resolución genética de los marcadores para poder detectar los procesos históricos, además de que las interpretaciones biológicas están limitadas por el tamaño de muestra y el número de sitios muestreados (Knowles y Maddison, 2002; Templeton, 2004).

Filogeografía de vertebrados

Los primeros estudios que utilizaron como herramienta el mtADN y que pueden ahora considerarse con una perspectiva filogeográfica explícita, fueron trabajos poblacionales de roedores, *Peromyscus* spp. y *Geomys pinetis* (Avise *et al.*, 1979a, b), basados en enzimas de restricción, cuya introducción abrió la posibilidad de estudiar la estructura genética a nivel del mtADN entre poblaciones de una especie a lo largo de su área de distribución. Así, Avise *et al.* (1979b) revisaron 87 individuos de *Geomys pinetis* en el sureste de Estados Unidos de América, encontrando con seis enzimas de restricción 23 haplotipos diferentes con los que estimaron la divergencia entre "clones" o grupos de mtADN y, como el aspecto más innovador del estudio, la filogenia intraespecífica matriarcal de la especie.

Considerando exclusivamente la información genética y mientras viajaba en un tren, Avise dibujó sobre un pedazo de servilleta la filogenia que interconectaba los 23 haplotipos, tomando en cuenta el número mínimo inferido de pasos mutacionales. Los aspectos filogeográficos del análisis cristalizaron cuando dicha filogenia se sobrepuso a la distribución geográfica de los sitios de muestreo, lo que reveló una marcada localización de los haplotipos y un componente histórico evidente de las relaciones matrilineales, incluyendo la separación entre todos los individuos de una porción este y otra porción oeste del área de distribución. Así se hacía evidente que cada individuo mostraba un haplotipo especificable, heredado intacto, sin recombinación genética intermolecular a través de los ancestros maternos, y con ello se sentaban las bases para lo que más tarde fue reconocido como la filogeografía.

A partir del trabajo de *Geomys* se desarrollaron múltiples estudios bajo esta perspectiva, lo que permitió ocho años más tarde hacer una revisión extensa de evidencias y resultados que Avise y colaboradores conjuntaron para definir cinco patrones filogeográficos particulares (ver Avise *et al.*, 1987; ver capítulo 15). Durante este tiempo, otro aspecto que resulta relevante es la importancia que tuvieron (y tienen) los vertebrados en general y los mamíferos en parti-

cular en el desarrollo de la filogeografía, por la mayor información geográfica y de distribución que se tenía comparada con otros grupos de animales. A partir de la década de los 90 se ha acumulado gran cantidad de estudios filogeográficos que involucran muchos taxones, desde invertebrados hasta plantas superiores, así como diferentes marcadores y técnicas moleculares. Asimismo, los estudios abarcan preguntas cada vez más específicas –no sólo de distribución geográfica– relativas a aspectos de hibridización, refugios pleistocénicos, sistemática, conducta, migración y conservación ¡e incluso sobre la polémica del concepto biológico y filogenético de especie! (véase el capítulo 10 de este libro).

Retomando la importancia de los vertebrados es indiscutible que éstos siguen siendo los más estudiados bajo perspectivas filogeográficas, comparativamente con plantas e invertebrados. Para desglosar esto con más detalle, hice una revisión de todos los trabajos en los que la palabra filogeografía aparece en el título o en las palabras clave, entre los años 2000 y 2003, en vertebrados. Para el periodo revisado, que refleja mucho de lo que se ha venido trabajando en los últimos años es evidente que hay taxones más estudiados, como los mamíferos y los peces, y en particular roedores y peces de agua dulce (figura 3).

Hay también un incremento notable del uso de marcadores nucleares, sobre todo microsatélites, aunque algunos estudios utilizan aloenzimas, comúnmente en combinación con mitocondriales y en menor medida por sí solos. Unos pocos han empezado a explorar el uso de caracteres morfológicos, aunque reconociendo las limitaciones que conllevan. Aunque la región control y el citocromo-*b* siguen siendo los genes mitocondriales preferidos, cada vez más trabajos emplean otros genes mitocondriales, como el 12 y 16S, NDH y ATP. Por otro lado, si se desglosa la información por país (dónde está hecho el estudio) y por taxon, en todos se observa un predominio de mamíferos y peces, con un incremento reciente perceptible para reptiles. Las aves, seguramente por su alta vagilidad y mayor dificultad en la delimitación de sus áreas geográficas se han mantenido con poca presencia en estos estudios (pero ver capítulo 15). Y no es Estados Unidos de América donde más se ha trabajado, pues Europa tiene una historia realmente importante en cuanto a los trabajos filogeográficos de sus vertebrados. Tal vez la diferencia más evidente entre estas regiones es que en Europa predominan un poco más los estudios de peces, mientras que en Estados Unidos de América, sobre todo recientemente, hay más trabajos con reptiles y anfibios (figura 4).

Figura 3. Número de artículos publicados entre 2000 y 2003, sobre filogeografía para cada grupo taxonómico de vertebrados

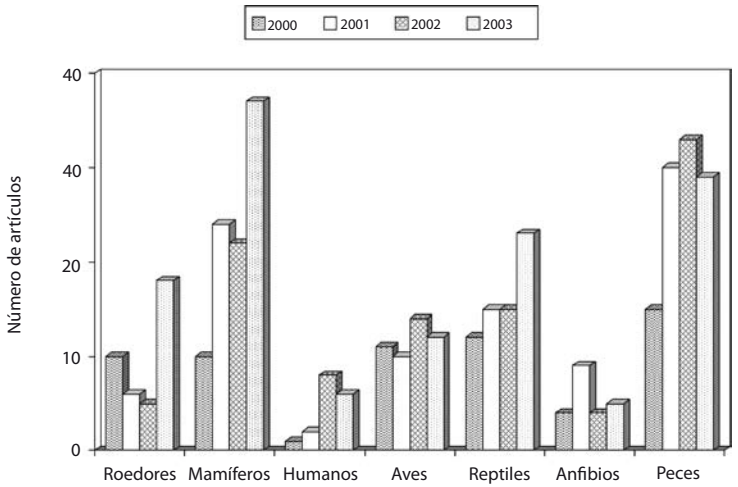
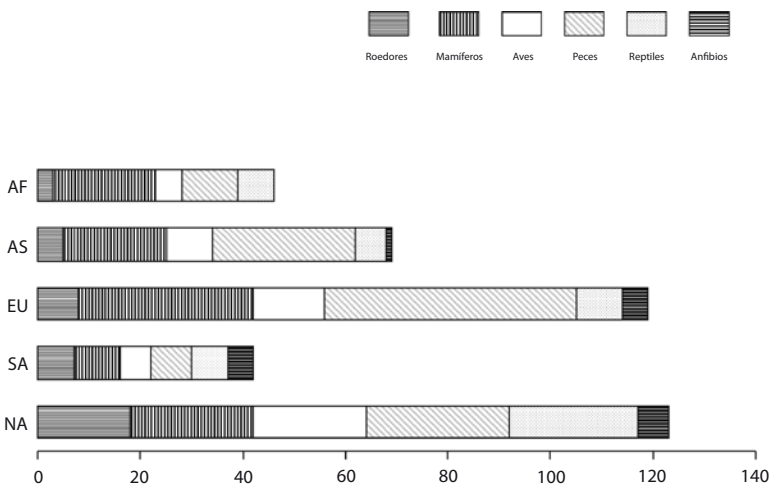


Figura 4. Proporción de artículos publicados en diferentes regiones (AF: África, AS: Asia, EU: Europa, SA: Sudamérica, NA: Norteamérica, sin incluir México) sobre filogeografía y para cada grupo taxonómico



De los muy diversos estudios filogeográficos con vertebrados, se pueden describir sólo unos pocos patrones generales, aunque para cada grupo existen excepciones y un sinnúmero de posibilidades dentro del espectro que se describe:

Mamíferos

Presentan una gran variedad de patrones; las especies pequeñas y poco vágiles tienen diferenciación filogeográfica profunda, mientras los vágiles, grandes, cuentan típicamente con estructuras menos pronunciadas. Los mamíferos marinos y voladores, aunque con alta movilidad, exhiben cierta estructura, relacionada principalmente con restricciones de dispersión, fidelidad al sitio o por la conformación del grupo social.

Aves

Se observan poblaciones coespecíficas bien estructuradas, con separación profunda de linajes que concuerda con diferencias en otros caracteres (conducta, morfología, área biogeográfica), hasta separaciones someras por fragmentación reciente y limitado flujo génico (véase el capítulo 15 de este libro).

Herpetofauna

La mayoría de los anfibios y reptiles presentan diferenciación filogeográfica profunda, básicamente por su poca vagilidad. Excepción son las tortugas marinas, con gran dispersión, que tienen estructuras tanto profundas (eventos vicariantes históricos y dispersión) como someras (fidelidad al sitio).

Peces

Con diferencias notables entre peces dulceacuícolas y marinos, con estructura pronunciada los primeros, en áreas pequeñas, hasta con distribuciones vastas y diferentes niveles de estructuración los segundos. Ello íntimamente relacionado con las diferencias en cuanto a hábitat y características físicas de cada ambiente.

Filogeografía intraespecífica

El estudio de la distribución geográfica de linajes de genes de poblaciones de una misma especie se define como filogeografía intraespecífica. Es bajo esta perspectiva que se han hecho la gran mayoría de los estudios filogeográficos: 75% del total de estudios entre 2000 y 2003; el resto, 25% fueron estudios de filogeografía comparada.

El inicio y *Homo sapiens*

Los estudios basados en mtADN que buscaban reconstruir las relaciones filogenéticas del hombre, corroborar su origen africano y determinar las diferentes migraciones hacia Europa y Asia, son reconocidos como estudios pioneros que permitieron además refinar la metodología empleada y que ésta se extendiera al resto de los grupos animales.

En sus inicios, en un trabajo de Cann *et al.* (1987; en Avise, 2000) se demostró con análisis de secuencias de mtADN el origen africano del linaje humano, que presentaba baja divergencia entre haplotipos y una limitada estructura filogeográfica. En uno de los estudios más recientes, el enfoque filogeográfico permitió determinar rutas de colonización y niveles de estructuración poblacional de *H. sapiens* a lo largo de ¡1.7 millones de años! (Templeton, 2002). El trabajo incluyó el uso de diferentes genes mitocondriales y nucleares, así como análisis de clados anidados (NCA), corroborando nuevamente el origen africano y demostrando también que existieron dos migraciones más de África a Asia, así como una más reciente en sentido contrario (de Asia a África) y que hubo cierto flujo génico constante entre las poblaciones, limitado por distancia geográfica. También se registró con certeza la expansión de poblaciones asiáticas a América.

La identificación de eventos de migración realmente recientes también es posible: Carvalho-Silva *et al.* (2001) examinaron la porción no recombinante del cromosoma Y en individuos de Brasil y Portugal, lo compararon con datos de mtADN, y pudieron probar que el origen (patrilineal) de los brasileños es significativamente europeo, mientras que el componente matrilineal tenía un importante componente de genes de origen africano e indoamericano. Pudieron detallar además los eventos de migración ocurridos alrededor de los 1500 d.C., siendo éstos de la región italiana hacia el sur de Brasil, y de Marruecos y Holanda a la región amazónica y al norte, respectivamente.

Otros vertebrados y diversas aplicaciones de la filogeografía

Una de las aplicaciones más amplias de los estudios filogeográficos ha sido el poder determinar el grado de estructuración poblacional de las especies a lo largo de su área de distribución, así como descifrar cuáles han sido los procesos que han determinado dicha distribución. El patrón filogeográfico más conspicuo es aquel en el que hay una marcada subdivisión genealógica entre poblaciones o grupos de poblaciones (haplogrupos) y una marcada estructuración espacial (alopátrica); generalmente se observa en escalas regionales amplias y es un patrón común en pequeños mamíferos, insectos y herpetofauna, dada su poca vagilidad.

Por ejemplo, el conejo rojo de Smith (*Pronolagus rupestris*) presenta en el sur de África dos haplogrupos altamente diferenciados, separados por más de 40 pasos mutacionales, uno del oeste y otro del sureste (Matthee y Robinson, 1996). Comúnmente, las separaciones mayores en los filogrupos coinciden con evidencia independiente (geológica, morfológica, de cariotipo, etc.), que muestra el aislamiento de las poblaciones por periodos largos, así como divergencia.

Aún cuando la mayoría de los haplotipos presentes en una especie estén cercanamente relacionados, se ha podido observar cierta estructuración geográfica; tal es el caso del roedor *Peromyscus polionotus*, en el sureste de Estados Unidos de América, que presenta un flujo génico contemporáneo que ha sido suficientemente bajo para permitir divergencia genética entre poblaciones, aunque las poblaciones han estado en contacto histórico reciente. Ello contrasta con lo encontrado, utilizando el mismo análisis, en *Peromyscus maniculatus* que habita el resto de Estados Unidos de América y presenta alta estructuración geográfica. La información en conjunto refleja una probable separación reciente de *P. polionotus* del grupo ancestral durante la dispersión de la especie hacia el sureste (Avise *et al.*, 1983).

Los peces de agua dulce en general presentan patrones filogeográficos marcados, sobre todo aquellos confinados a cuerpos de agua aislados como el pez de desierto *Gila atraria*, en el que se observan dos filogrupos distintivos, asociados a un río y a un lago; dentro de los filogrupos hay a su vez estructuración de las poblaciones, asociada a eventos vicariantes concordantes con la historia hidrológica pleistocénica de la zona (Johnson, 2002).

Por el contrario, se tienen ejemplos de especies prácticamente sin ninguna estructuración poblacional, como las especies marinas (Avise, 2000). *Cleve-*

landia ios y *Eucyclogobius newberryi* son dos peces marinos simpátricos de California, el primero con una estructura filogeográfica considerablemente menor comparada con la del segundo, lo que está directamente relacionado con la mayor habilidad de dispersión de *C. ios* (Dawson *et al.*, 2002).

Por su lado, el gorrión *Ammodramus savannarum* es muy peculiar: Bulgin *et al.* (2003) estudiaron la región control y seis microsatélites de cinco poblaciones de esta ave, distribuidas en diferentes estados de Norteamérica, una de las cuales está aislada (estado de Florida) y representa una subespecie en peligro; contrario a lo esperado no encontraron estructuración alguna. Por medio de análisis de coalescencia pudieron determinar que el tamaño efectivo femenino de las poblaciones es alto (~ 80000 hembras) y que divergieron recientemente (< 26 000 años); por lo tanto, la falta de diferenciación filogenética de la subespecie y la mínima diferenciación genética se debe a la retención de polimorfismos ancestrales, que enmascaran la detección de poblaciones evolutivamente diferentes (véase el capítulo 3 de este libro).

Entre los mamíferos de talla grande, a pesar de que se esperaría poca estructuración por su alta movilidad, se han encontrado resultados contrastantes. En el chimpancé, *Pan troglodytes*, la dispersión de individuos es alta (600-900 km), por lo que hay alto flujo génico y nula diferenciación morfológica entre 3 subespecies africanas. Sin embargo, dos haplotipos marcadamente divergentes distinguen a *Pan troglodytes verus*, subespecie bien diferenciada y evolutivamente independiente (Morin *et al.*, 1994). Por su parte, el leopardo *Panthera pardus* presenta seis filogrupos geográficamente aislados, definidos con base en mtADN y microsatélites, y que demuestran el origen africano y posterior migración a Asia de la especie (Uphyrkina *et al.*, 2001), pero estos filogrupos no coinciden con las subespecies que se reconocen taxonómicamente.

Irwin (2002) demostró que existen casos en los que puede presentarse una marcada separación filogeográfica aun sin barreras al flujo génico. Propone dos explicaciones para el caso específico del ave *Phylloscopus trochiloides*, que presenta dos filogrupos (este y oeste) dentro de su área de distribución en Siberia: que la especie ancestral fue separada en dos por una barrera geográfica, posteriormente las poblaciones se expandieron y ahora están en contacto (dos filogrupos, sin barrera y con cierto flujo génico), o que la especie ha estado distribuida siempre en toda el área, y que la diferenciación en mtADN es resultado de la baja dispersión individual.

Un caso diferente con otra ave de Norteamérica, *Wilsonia pusilla*, es el que reportan Kimura *et al.* (2002). Encontraron diversos grados de estructuración: la especie se divide en dos filogrupos marcados (este *vs* oeste), pero dentro de

cada uno la estructuración es distinta; existe cierta estructuración dentro del grupo del oeste, pero es prácticamente nula en el este. Ello estaría relacionado con las altas tasas de migración a los sitios de reproducción y con niveles más bajos de migración a los sitios de alimentación, ya que estos eventos se dan en diferentes épocas del año y en diferente intensidad en las poblaciones del este y del oeste. Este ejemplo muestra que los patrones filogeográficos pueden ser sensibles a procesos actuales (y no sólo históricos), como la migración estacional, así como a la conducta y estructura social de las especies. Otro ejemplo es el del primate *Callithrix jacchus*, en el que se encontró estructuración significativa de haplotipos tanto entre grupos sociales (colonias) como entre poblaciones (sitios). Sorpresivamente, el 85% de las parejas reproductivas presentan linajes maternos diferentes, lo que demuestra entrecruzas aleatorias y exogamia, lo que es poco frecuente en especies altamente sociales (Faulkes *et al.*, 2003).

Las interpretaciones históricas de las unidades filogeográficas de mtADN están comúnmente apoyadas por evidencia geográfica, tal es el caso de especies que presentan linajes con una distribución consistente con la teoría de refugios del Pleistoceno. Muchos trabajos se han concentrado en evaluar el efecto de las fluctuaciones climáticas del Plioceno y Pleistoceno en la distribución actual de la biodiversidad, y entre otros hallazgos se ha podido comprobar que diversos eventos de especiación en vertebrados que tradicionalmente se habían determinado para el Pleistoceno, son en realidad mucho más antiguos (ver Avise y Walker, 1998). Utilizando herramientas filogeográficas (gen 16S rARN y análisis de clados anidados) Veith *et al.* (2003) pudieron inferir los escenarios de expansión del área de distribución, así como los de fragmentación que determinaron la distribución actual de cinco especies de ranas de Anatolia (Turquía): la amplia distribución original del ancestro común en Anatolia se fragmentó en diversas poblaciones durante las glaciaciones. Como resultado de sucesivos eventos de vicarianza y dispersión a lo largo del Plioceno y Pleistoceno, se formaron poblaciones relictas restringidas a refugios, que representan actualmente centros de endemismo a nivel de haplotipos, congruentes con los tres linajes reconocidos (*Rana macrocnemis*, *R. camerani*, *R. holtzi*).

Otro ejemplo son los eventos climáticos y geológicos del Cuaternario tardío, que fueron determinantes en la estructuración de comunidades bióticas del norte del hemisferio boreal, sobre todo en la región conocida como Beringia o estrecho de Bering. Diversos taxones divergieron entre Europa, Asia y Norteamérica, como resultado de las migraciones y periodos de aislamiento

durante los ciclos glaciares e interglaciares, así como de las barreras geográficas presentes en la zona. Estudios filogeográficos del roedor *Microtus oeconomus* muestran cuatro filogrupos bien diferenciados a lo largo de su distribución asociados con eventos vicariantes: norte y centro de Europa, separados éstos a su vez del centro de Asia por los montes Urales y por último el clado de Beringia, conformado por el este de Rusia, Alaska y noroeste de Canadá (Brunhoff *et al.*, 2003). Patrones similares se han observado en lemures (Fedorov *et al.*, 2003), mientras que en pequeños mamíferos de Amazonía oriental, por el contrario, se encontró que los cambios pleistocénicos no ocasionaron expansiones demográficas (Lessa *et al.*, 2003). En el género *Lepus* (conejos y liebres), con base en la filogeografía se pudo determinar su origen, evolución y eventos de colonización y recolonización. Se confirmó que el género se originó en América, particularmente a partir de *Lepus americanus*; mientras que las distancias genéticas observadas y los patrones de dispersión estimados revelaron eventos de migración congruentes con la historia geográfica de la zona, que van inicialmente de América hacia Asia (por el estrecho de Bering), de donde continúan hacia Europa, para finalmente regresar a la región holártica a través de Asia (Ramírez-Silva *et al.* en revisión).

Gracias al análisis de ADN, se han podido también detallar eventos de extinción precisos, como el caso de *Ursus arctos* (oso café), con base en el análisis de ADN preservado en restos animales atrapados en capas de hielo, de hasta 60 000 años de antigüedad (Barnes *et al.*, 2002).

Otra aplicación relevante de los estudios filogeográficos tiene que ver con la posibilidad de probar diferentes modelos de diversificación. Schneider *et al.* (1999) compararon la divergencia mitocondrial y morfológica de ocho poblaciones de la lagartija *Carlia rubrigularis*. Encontraron que las diferencias morfológicas entre poblaciones geográficamente aisladas (separadas por varios millones de años, de acuerdo con el mtADN) eran mínimas, mientras que las diferencias entre poblaciones de distintos hábitats eran mayúsculas. Con ello pudieron probar que la selección natural, operando en gradientes ecológicos, es más importante que el aislamiento geográfico (refugios) en la diversidad fenotípica (y probable especiación) de esta lagartija.

Los aspectos de sistemática y taxonomía han sido también frecuentemente abordados a través de la filogeografía (como se muestra en varios de los ejemplos ya mencionados). El jaguar, *Panthera onca*, no presenta una estructura geográfica marcada en sus poblaciones que van desde el centro de México hasta el sur de Sudamérica, lo que se atribuye a una expansión demográfica reciente y poco tiempo de aislamiento (Eizirik *et al.*, 2001). Dichos resultados

contrastan con las subespecies propuestas hasta ahora, por lo que los autores proponen una reconsideración taxonómica completa de la especie, sobre todo dada su categoría de especie en riesgo (véase el capítulo 8 de este libro). Asimismo, la presencia de barreras paleogeográficas entre Amazonía del este y del oeste ha generado divergencia genética significativa en muchas especies de vertebrados, entre los cuales está el género *Dendrobates* (rana venenosa), y para el que se propone una revisión taxonómica completa con base en los resultados filogeográficos obtenidos con cuatro genes mitocondriales (Symula *et al.*, 2002).

En México existen 10 especies del roedor *Cratogeomys*, la mayoría restringidas al cinturón trans-volcánico. En este grupo, Demastes *et al.* (2002) encontraron cinco clados que concuerdan con la geografía, edafología y la evolución del cinturón en el periodo Cuaternario; en ellos, la distribución de cuatro especies refleja adecuadamente su taxonomía, pero sugieren que el resto del grupo requiere revisión. También se ha propuesto revisar la taxonomía de *Neotoma mexicana*, especie de roedor con distribución en Estados Unidos de América y México, dados los resultados de un estudio filobiogeográfico (*sensu* Edwards y Bradley, 2002). Pero, finalmente, es importante mencionar que un estudio filogeográfico es sólo uno de los muchos aspectos que se requieren como información definitoria para determinar cambios taxonómicos de especies (véase los capítulos 10 y 11 de este libro).

En otra especie de roedor mexicano, *Peromyscus furvus*, de la Sierra Madre Oriental, se identificaron cinco filogrupos con altos niveles de diferenciación genética basados en mtADN y aloenzimas. Dada la posición en el filograma y la posición geográfica del clado más sureño, compuesto por una sola población que se estimó divergió hace 1.3 a 1.8 ma, se sugiere que pueda ser una especie diferente (Harris *et al.*, 2000). Por otro lado, como todos los filogrupos tienen distribución alopátrica restringida a las cimas de la Sierra, los autores las proponen como *Unidades Significativamente Evolutivas*. Ello nos lleva a otra de las aplicaciones comúnmente utilizadas de la información filogeográfica: la conservación de la biodiversidad. Las Unidades Significativamente Evolutivas (ESUs por sus siglas en inglés; Moritz, 1995), representan grupos de poblaciones que han estado históricamente aisladas y que en conjunto representan la diversidad de un taxón (véase el capítulo 8 de este libro).

En este sentido la fauna de las islas oceánicas ha sido fuente de estudios genéticos y de conservación. Las tortugas gigantes de las Galápagos *Geochelone nigra*, comprenden un linaje que radió de manera rápida junto con la evolución del archipiélago, historia que se ha podido detallar utilizando mtA-

DN y microsatélites. Las poblaciones tienen marcada divergencia genética, historias demográficas contrastantes y estructura filogeográfica profunda, consistente con la historia geológica y biogeográfica de las islas, por lo que se han propuesto cuatro ESUs (dos en Santa Cruz, una en Pinzón y una en San Cristóbal; Beheregaray *et al.*, 2003). Por su parte, el elefante asiático (*Elephas maximus*) presenta poblaciones muy reducidas y fragmentadas, tanto por factores históricos como por dispersión reciente mediada por el hombre. Un estudio filogeográfico muestra dos clados que se separaron hace cerca de 1.2 ma. Los individuos de los dos clados están presentes en todas las localidades de estudio, excepto Indonesia y Malasia, que están en linajes basales, por lo que se les confiere estatus de ESUs (Fleischer *et al.*, 2001).

Filogeografía comparada

La filogeografía comparada estudia los efectos de la historia y la biogeografía en la distribución de la variación genética de varias especies de una comunidad, esto es, son estudios filogeográficos regionales que consideran los patrones espaciales de los linajes de múltiples taxones codistribuidos. La resolución de este tipo de estudios es mayor, puesto que permite evaluar las asociaciones entre procesos poblacionales de diferentes taxones que han evolucionado independientemente, pero en las que los patrones regionales filogenéticos y geográficos son concordantes, lo que se considera como evidencia de la influencia de factores ecológicos y/o evolutivos en la estructura y la distribución actual de las especies. La diversidad de especies y taxones que puede evaluarse en términos comparativos es amplísima, como se verá en los ejemplos siguientes.

Pueden compararse linajes diferentes de una misma especie, como en el salmón rosado, *Oncorhynchus gorbuscha* que tiene un ciclo de vida estricto de dos años. Esta especie tiene tiene dos poblaciones reproductivamente aisladas en el norte del Océano Pacífico, cada una de las cuales se reproduce en años diferentes (alternados). Utilizando mtADN y NCA se encontraron cinco filogrupos, y dentro de cada uno, una filogenia en estrella que indica crecimiento poblacional reciente, lo que puede ser resultado de la expansión del rango histórico junto con episodios de dispersión restringida (Churikov y Gharrett, 2002). Otro estudio con peces, del complejo *Odontesthes perugiae* de las costas de Brasil (mtADN y microsatélites), apoya la propuesta de un origen marino-estuarino del complejo, y una radiación cronológicamente asociada a los cambios del nivel del mar del Pleistoceno y Holoceno, que se ha traducido en tres especies alopátricas y dos simpátricas (Beheregaray *et al.*, 2002).

Ditchfield (2000) comparó los patrones filogeográficos de murciélagos y pequeños mamíferos no voladores neotropicales (del centro de México al sur de Sudamérica). Los murciélagos mostraron poca estructuración, con clados formados por haplotipos muy similares presentes prácticamente a todo lo largo de la distribución. Por el contrario, los mamíferos no voladores consistentemente presentaron niveles altos de divergencia, en distancias geográficas pequeñas. Estos resultados no sólo reflejan las diferencias biológicas de las especies, sino también que el efecto geográfico varía dependiendo de la escala. La estructura filogeográfica de dos especies de aves en seis montañas de Camerún y Guinea Ecuatorial muestra dos patrones: uno concordante en las dos especies que permite definir tres regiones biogeográficas relevantes (que divergieron por alopatría y no sólo por aislamiento por distancia); y otro en el que dentro de cada especie se observan linajes monofiléticos (endemismos) para dos de las tres regiones (Smith *et al.*, 2000). Con lo anterior se aprecia la ventaja de la filogeografía comparada para identificar regiones que albergan poblaciones evolutivamente distintas con valor de conservación.

Combinando taxones, Evans *et al.* (2003) pudieron definir áreas de endemismo. Estudiaron diversas especies de monos (*Macaca* spp.) y sapos (*Bufo* spp.) endémicos de la isla de Sulawesi (Indonesia) y encontraron que la distribución a una escala geográfica fina de ambos grupos era virtualmente idéntica, y en ella se aprecian siete áreas multi-taxa de endemismo genético. Los patrones concordantes de diversidad pueden así facilitar estrategias de conservación, dado que pueden designarse áreas de protección estratégicas que contengan componentes de variación complementarios y no redundantes (Evans *et al.*, 2003). En uno de los estudios comparados más completos, Riddle *et al.* (2000) analizaron la estructura filogeográfica de 12 especies de mamíferos, aves, anfibios y reptiles con dos genes mitocondriales. Con la comparación de la distribución de los filogrupos, se corroboraron eventos vicariantes propuestos para la formación de la península de Baja California (México) durante el Pleistoceno. Esta historia evolutiva particular del desierto peninsular no había podido describirse, ya que se había estudiado con especies de amplia distribución. La filogeografía comparada permitió determinar que la península ya no puede considerarse un subgrupo del desierto de Sonora, y que tiene su propia historia evolutiva y ecológica, así como su propio valor de conservación (Riddle *et al.*, 2000).

La filogeografía comparada también ha contribuido a mejorar los métodos cuantitativos aplicados a la selección de áreas para protección que maximi-

cen la representación no sólo de riqueza y diversidad de especies, sino que incorporen explícitamente su persistencia, *i.e.* que protejan los procesos que sostienen dicha diversidad (Ferrier, 2002). Analizando los patrones filogeográficos comparados de reptiles, aves y anfibios de la selva tropical australiana, que representan patrones de aislamiento histórico y los procesos evolutivos asociados, se pudo hacer la selección de áreas prioritarias para conservación, que acumularan la mayor divergencia genética. Se corroboró que dicha selección de áreas prioritarias era más eficiente (acumulaba la mayor diversidad en el menor número de áreas), comparativamente con una selección basada en número de especies. Como puede apreciarse con los diversos estudios descritos, la información filogeográfica y de diversidad de especies puede combinarse para desarrollar estrategias de conservación a nivel de comunidades y de subregiones (Vázquez-Domínguez, 2002).

Como reflexiones finales, podemos resaltar que los mamíferos han sido y siguen siendo un grupo primordial de estudio en la filogeografía, aunque cada vez se involucran más otros grupos, al igual que existe una gran diversidad de preguntas que pueden abordarse bajo la perspectiva filogeográfica, y que aunque existen temas más trabajados, el campo de estudio es prácticamente inagotable. Las herramientas y métodos de análisis filogeográfico están en pleno desarrollo, y es seguro que su perfeccionamiento y alcance serán mayores en el corto plazo. Finalmente, como pudo apreciarse a lo largo del presente capítulo, los ejemplos existentes para el caso de vertebrados mexicanos son mínimos: en particular, existen no más de 16 estudios entre 2000 y 2003, de los cuales sólo cuatro son hechos exclusivamente con taxones en México, el resto involucra especies que tienen sólo una parte de su distribución en el país (Vázquez-Domínguez y Hafner, 2002). Esto simplemente significa que la filogeografía es un área de estudio que no ha sido desarrollada en su plenitud en México, a pesar de la intrincada y extraordinaria historia geológica y biogeográfica de buena parte del territorio, que permitiría abordar gran diversidad de preguntas evolutivas sobre muchas especies de fauna y flora.

Bibliografía

Avise, J.C. 2000. *Phylogeography, the history and formation of species*. Harvard University Press, Cambridge, EUA.

Avise, J.C. y D. Walker. 1998. Pleistocene phylogeographic effects on avian populations and the speciation process. *Proceedings of the Royal Society of London Series B* 265:457-463.

Avise, J.C., R.A. Lansman y R.O. Shade. 1979a. The use of restriction endonucleases to measure mitochondrial DNA sequence relatedness in natural populations I. Population structure and evolution in the genus *Peromyscus*. *Genetics* 92:279-295.

Avise, J.C., C. Giblin-Davidson, J. Laerm, J.C. Patton y R.A. Lansman. 1979b. Mitochondrial DNA clones and matriarchal phylogeny within and among geographic populations of the pocket gopher, *Geomys pinetis*. *Proceedings of the National Academy of Sciences* 76:6694-6698.

Avise, J.C., J.F. Shapira, S.W. Daniel, C.F. Aquadro y R.A. Lansman. 1983. Mitochondrial DNA differentiation during the speciation process in *Peromyscus*. *Molecular Biology and Evolution* 1:38-56.

Avise, J.C., J. Arnold, R.M. Ball, E. Bermingham, T. Lamb, J.E. Niegel, C.A. Reeb y N.C. Saunders. 1987. Intraspecific phylogeography: the mitochondrial DNA bridge between population genetics and systematics. *Annual Review of Ecology and Systematics* 18:489-522.

Barnes, I., P. Matheus, B. Shapiro, D. Jensen y A. Cooper. 2002. Dynamics of Pleistocene population extinctions in Beringian brown bears. *Science* 295:2267-2270.

Beheregaray, L.B., P. Sunnucks y D.A. Briscoe. 2002. A rapid fish radiation associated with the last sea-level changes in southern Brazil: the silverside *Odontesthes perugiae* complex. *Proceedings of the Royal Society of London Series B* 269:65-73.

Beheregaray, L.B., C. Ciofi, A. Caccone, J.P. Gibbs y J.R. Powell. 2003. Genetic divergence, phylogeography and conservation units of giant tortoises from Santa Cruz and Pinzon, Galapagos Islands. *Conservation Genetics* 4:31-46.

Brown, W.M. 1985. The mitochondrial genome of animals. En: R.J. MacIntyre (ed.). *Molecular evolutionary genetics*. Plenum, Nueva York, EUA. pp. 95-130.

Brown, W.M., M. George Jr. y A.C. Wilson. 1979. Rapid evolution of animal mitochondrial DNA. *Proceedings of the National Academy of Sciences* 76:1967-1971.

Brunhoff, C., K.E. Galbreath, V.B. Fedorov, J.A. Cook y M. Jaarola. 2003. Holartic phylogeography of the root vole (*Microtus oeconomus*): implications for late Quaternary biogeography of high latitudes. *Molecular Ecology* 12:957-968.

Bulgin, N.L., H.L. Gibbs, P. Vickery y A.J. Baker. 2003. Ancestral polymorphisms in genetic markers obscure detection of evolutionary distinct populations in the endangered Florida grasshopper sparrow (*Ammodramus savannarum floridanus*). *Molecular Ecology* 12:831-844.

Carbone, I. y L.M. Kohn. 2001. A microbial population-species interface: nested cladistics and coalescent inference with multilocus data. *Molecular Ecology* 10:947-964.

Carvalho, G.R. 1998. Molecular ecology: origins and approach, en G.R. Carvalho (ed.) Advances in molecular ecology. IOS Press, Oxford, Inglaterra. págs. 1-23.

Carvalho-Silva, D.R., F.R. Santos, J. Rocha y S.D.J. Peña. 2001. The phylogeography of Brazilian Y-chromosome lineages. *American Journal of Human Genetics* 68:281-286.

Churikov, D. y A.J. Gharrett. 2002. Comparative phylogeography of the two pink salmon broodlines: an analysis based on a mitochondrial DNA genealogy. *Molecular Ecology* 11:1077-1101.

Cruzan, M.B. y A.R. Templeton. 2000. Paleoecology and coalescence: phylogeographic analysis of hypotheses from the fossil record. *Trends in Ecology and Evolution* 15:491-496.

Dawson, M.N., K.D. Louie, M. Barlow, D.K. Jacobs y C.C. Swift. 2002. Comparative phylogeography of sympatric sister species, *Clevelandia ios* and *Eucyclogobius newberryi* (Teleostei, Gobiidae), across the California Transition Zone. *Molecular Ecology* 11:1065-1075.

Demastes, J.W., T.A. Spradling, M.S. Hafner, D.J. Hafner y D.L. Reed. 2002. Systematics and phylogeography of pocket gophers in the genera *Cratogeomys* and *Pappogeomys*. *Molecular Phylogenetics and Evolution* 22:44-154.

Ditchfield, A.D. 2000. The comparative phylogeography of Neotropical mammals: patterns of intraspecific mitochondrial DNA variation among bats contrasted to nonvolant small mammals. *Molecular Ecology* 9:1307-1318.

Edwards, C.W. y R.D. Bradley. 2002. Molecular systematics and historical phylobiogeography of the *Neotoma mexicana* species group. *Journal of Mammalogy* 83:20-30.

Eizirik, E., J.H. Kim, M. Menotti-Raymond, P.G. Crawshaw, S.J. O'Brien y W.E. Johnson. 2001. Phylogeography, population history and conservation genetics of jaguars (*Panthera onca*, Mammalia, Felidae). *Molecular Ecology* 10:65-79.

Evans, B.J., J. Supriatna, N. Andayani y D.J. Melnick. 2003. Diversification of Sulawesi macaque monkeys: decoupled evolution of mitochondrial and autosomal DNA. *Evolution* 57:1931-1946.

Faulkes, C.G., M.F. Arruda y A.O.M. Da Cruz. 2003. Matrilineal genetic structure within and among populations of the cooperatively breeding common marmoset, *Callithrix jacchus*. *Molecular Ecology* 12:1101-1108.

Fedorov, V.B., A.V. Goropashnaya. M. Jaarola y J.A. Cook. 2003. Phylogeography of lemmings (*Lemmus*): no evidence for postglacial colonization of Arctic from the Beringian refugium. *Molecular Ecology* 12:725-731.

Ferrier, S. 2002. Mapping spatial pattern in biodiversity for regional conservation planning: where to from here? *Systematic Biology* 51:331-363.

Fleischer, R.C., E.A. Perry, K. Muralidharan, E.E. Stevens y C.M. Wemmer. 2001. Phylogeography of the Asian elephant (*Elephas maximus*) based on mitochondrial DNA. *Evolution* 55:1882-1892.

Gould, S.J. 1980. Is a new and general theory of evolution emerging? Paleobiology 6: 119-130.

Harding, R.M. 1996. New phylogenies: an introductory look at the coalescent. En: P.H. Harey, A.J. Leigh Brown, J. Maynard Smith y S. Nee (eds.). *New uses for new phylogenies*. Oxford University Press, Nueva York, EUA. pp 15-22.

Hare, M.P. 2001. Prospects for nuclear gene phylogeography. *Trends in Ecology and Evolution* 16:700-706.

Harris, D, D.S. Rogers y J. Sullivan. 2000. Phylogeography of *Peromyscus furvus* (Rodentia; Muridae) based on cytochrome *b* sequence data. *Molecular Ecology* 9:2129-2135.

Hudson, R.R. 1990. Gene genealogies and the coalescent process. En: D. Futuyma, J. Antonovics (eds.). *Oxford surveys in evolutionary biology*. Oxford University Press, Oxford, Reino Unido. pp 1-44.

Hudson, R.R. 1998. Island models and the coalescent process. Molecular Ecology 7: 413-418.

Hudson, R.R., M. Slatkin y W.P. Maddison. 1992. Estimation of levels of gene flow from DNA sequence data. *Genetics* 132:583-589.

Irwin, D.E. 2002. Phylogeographic breaks without geographic barriers to gene flow. *Evolution* 56:2383-2394.

Johnson, J.B. 2002. Evolution after the flood: Phylogeography of the desert fish Utah chub. *Evolution* 56:948-960.

Kimura, M., S. Clegg, I.J. Lovette, K.R. Holder, D.J. Girman, B. Milá, P. Wade y T.B. Smith. 2002. Phylogeographical approaches to assessing demographic connectivity between breeding and overwintering regions in a Neartic-Neotropical warbler (*Wilsonia pusilla*). *Molecular Ecology* 11:1605-1616.

Knowles, L.L. y W.P. Maddison. 2002. Statistical phylogeography. *Molecular Ecology* 11:2623-2635.

Lessa, E.P., J.A. Cook y J.L. Patton. 2003. Genetic footprints of demographic expansion in North America, but not Amazonia, during the Late Quaternary. *Proceedings of the National Academy of Sciences* 100:10331-10334.

Lunt, D.H., L.E. Whipple y B.C. Hyman. 1998. Mitochondrial DNA variable number tandem repeats (VNTRs): utility and problems in molecular ecology. *Molecular Ecology* 7:1441-1455.

Matthee, C.A. y T.J. Robinson. 1996. Mitochondrial DNA differentiation among geographical populations of *Pronolagus rupestris*, Smith's red rock rabbit (Mammalia, Lagomorpha). *Heredity* 76:514-523.

Mayr, E. 1982. Speciation and macroevolution. *Evolution* 36:1119-1132.

Morin, P.A., J.J. Moore, R. Chakraborty, L. Lin, J. Goodall y D.S. Woodruff. 1994. Kin selection, social structure, gene flow and the evolution of chimpanzees. *Science* 265:1193-1201.

Moritz, C. 1995. Uses of molecular phylogenies for conservation. *Philosophical Transactions of the Royal Society of London B* 349:113-118.

Niegel, J.E y J.C. Avise. 1993. Application of a random-walk model to geographic distributions of animal mitochondrial DNA variation. *Genetics* 135:1209-1220.

Nordborg, M. (2000) Coalescent theory. En: D.J. Balding, M.J. Bishop y C. Cannings (eds.). *Handbook of statistical genetics*. John Wiley & Sons, Winchester, Reino Unido. pp 1-37.

Ramírez-Silva, J.P., F. González, E. Vázquez-Domínguez, F. Cervantes. (enviado). Phylogeography and evolution of *Lepus* (Mammalia: Lagomorpha). *Journal of Mammalogy*.

Riddle, B.R., D.J. Hafner, L.F. Alexander y J.R. Jaeger. 2000. Cryptic vicariance in the historical assembly of a Baja California peninsular desert biota. *Proceedings of the National Academy of Sciences* 97:14438-14443.

Schneider, C.J., T.B. Smith, B. Larison y C. Moritz. 1999. A test of alternative models of diversification in tropical rainforests: ecological gradients vs. rainforest refugia. *Proceedings of the Natural Academia of Sciences* 96:13869-13873.

Smith, T.B., K. Holder, D. Girman, K. O'Keefe, B. Larison y T. Chan. 2000. Comparative avian phylogeography of Cameroon and Equatorial Guinea mountains: implications for conservation. *Molecular Ecology* 9:1505-1516.

Symula, R., R. Schulte y K. Summers. 2002. Molecular systematics and phylogeography of Amazonian poison frogs of the genus *Dendrobates*. *Molecular Phylogenetics and Evolution* 26:452-475.

Templeton, A.R. 1983. Phylogenetic inference from restriction endonuclease cleavage site maps with particular reference to the evolution of man and the apes. *Evolution* 37:221-244.

Templeton, A.R. 1998. Nested clade analyses of phylogeographic data: testing hypotheses about gene flow and population history. *Molecular Ecology* 7:381-397.

Templeton, A.R. 2002. Out of Africa again and again. *Nature* 416:45-51.

Templeton, A.R. 2004. Statistical phylogeography: methods of evaluating and minimizing inference errors. *Molecular Ecology* 13:789-809.

Templeton, A.R., E. Routman y C. Phillips. 1995. Separating population structure from population history: a cladistic analysis of the geographical distribution of mitochondrial DNA haplotypes in the Tiger Salamander, *Ambystoma tigrinum*. *Genetics* 140:767-782.

Uphyrkina, O., W. Johnson, H. Quigley, D. Miquelle, L. Marker, M. Bush, y S.J. O'Brien. 2001. Phylogenetics, genome diversity and the origin of modern leopard, *Panthera pardus*. *Molecular Ecology* 10:2617-2633.

Vázquez-Domínguez, E. (en prensa). ¿Cuál es el enfoque de la ecología molecular? Ejemplos con un ratón espinoso. En: G. Sánchez Rojas y A.E. Rojas Martínez (eds.). *Tópicos en sistemática, biogeografía, ecología y conservación de mamíferos*. CIB-UAEH, México. (en prensa).

———. 2002. Phylogeography, historical patterns and conservation of natural areas. En: G. Nelson, J.C. Day, L.M. Sportza, J. Loucky y C. Vásquez (eds.). *Protected areas and the regional planning imperative in North America*. University of Calgary Press, Calgary, Canadá. pp 369-378.

Vázquez-Domínguez, E.y D.J. Hafner. 2006. *Genética y mamíferos mexicanos. presente y futuro*. New Mexico Museum of History and Science Bulletin 32, Albuquerque, EE.UU.

Veith, M., J.F. Schmidtler, J. Kosuch, I. Baran y A. Seitz. 2003. Palaeoclimatic changes explain Anatolian mountain frog evolution: a test for alternating vicariance and dispersal events. *Molecular Ecology* 12:185-199.

Wainscoat, J. 1987. Out of the garden of Eden. *Science* 234:194-196.

Wright, S. 1931. Evolution in Mendelian populations. *Genetics* 16:97-159.

———. 1943. Isolation by distance. *Genetics* 28:114-138.

Capítulo 15

Filogeografía de aves mexicanas

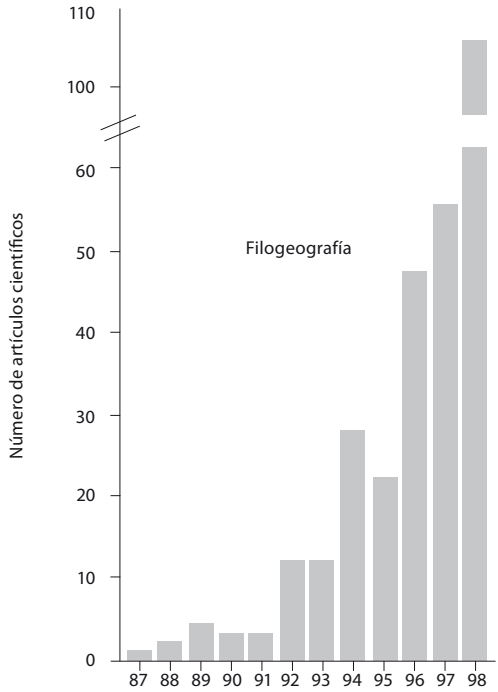
Blanca Estela Hernández-Baños,
Magali Honey-Escandón, María N. Cortés-Rodríguez
y Gabriela García Deras

La estructura genética poblacional de una especie es el resultado de los procesos microevolutivos y demográficos que actúan entre y dentro de las poblaciones, así como de la biogeografía histórica de los linajes genéticos (Wenink *et al.,* 1996). Debido a lo anterior, examinar las variaciones entre las poblaciones puede revelar sus asociaciones históricas y los procesos de reestructuración genética que pueden conducir a la especiación (Kirchman *et al.,* 2000). Los estudios de variación geográfica han revelado los mecanismos históricos y ecológicos, así como los patrones fenotípicos y de distribución de muchas especies de aves (Kirchman *et al.,* 2000), por lo que entender estos patrones de variación geográfica es el primer paso para inferir los procesos evolutivos de sus poblaciones naturales (Zink y Remsen, 1986).

La filogeografía es una disciplina reciente que surgió a raíz de los estudios de variación genética realizados con enzimas, RAPDs, RFLPs y el descubrimiento del ADN mitocondrial (mtADN) como molécula informativa para análisis filogenéticos y de genética de poblaciones (véase el capítulo 14 de este libro). La filogeografía es un campo que en últimas fechas ha tenido un gran desarrollo, observado principalmente en el número de publicaciones científicas que han utilizado la palabra filogeografía o filogeográfico (figura 1), y se han realizado numerosos estudios en diferentes reinos y phyla alrededor del mundo, incluidas las aves.

La filogeografía es la rama de la biogeografía que trata de los principios y procesos que rigen las distribuciones geográficas de los linajes genealógicos o filogenéticos, especialmente entre y dentro de las especies cercanamente relacionadas (Avise, 1998, 2000), es decir, tiene que ver con los componentes históricos y filogenéticos de las distribuciones espaciales de linajes de genes. El análisis de éstos y su interpretación requiere de datos provenientes de disciplinas como la genética molecular y de poblaciones, la etología, la demografía, la sistemática filogenética, la paleontología, la geología y la geografía histórica. Esto hace de la filogeografía una disciplina integrativa que se encuentra en el cruce entre la macro y la microevolución (Avise, 2000; véase también el capítulo 14 de este libro).

Figura 1. Aumento en el número de artículos científicos
que han utilizado a la filogeografía como herramienta de análisis
(Tomado de Avise, 2000)



A diferencia de los primeros análisis de genética de poblaciones realizados con aloenzimas (electroforesis de proteínas) en los que se reveló poca o ninguna diferenciación genética intraespecífica (Zink *et al*., 1991; Avise, 2000), los estudios realizados con ADN mitocondrial proporcionaron los primeros datos accesibles de genealogías de genes a nivel intraespecífico. Esto forjó un puente empírico y conceptual entre la genética de poblaciones (microevolución) y la filogenética tradicional (macroevolución; Avise, 1994). Haciendo una comparación de los árboles obtenidos a partir de genes mitocondriales y nucleares de varias especies de aves, se encontró que el ADN mitocondrial resolvía en un 95% la probabilidad de igualar la historia evolutiva de una especie, mientras que para el ADN nuclear solamente resolvía un 65%, siempre y cuando existiesen tamaños efectivos de población grandes al momento de la divergencia de estos genes. Con base en lo anterior, serían necesarios muchos

Tabla 1. Categorías de patrones filogeográficos y su descripción

Categoría I	Árbol genético con divergencia antigua, linajes alopátricos. Presencia de haplogrupos restringidos a un área geográfica determinada, separados por distancias genéticas relativamente grandes. La subestructura espacial también puede estar presente entre los linajes más cercanamente relacionados dentro de las regiones. Esta categoría comúnmente aparece en las investigaciones filogeográficas realizadas con mtADN.
Categoría II	Árbol genético con divergencia antigua, linajes ampliamente simpátricos. Se caracteriza por diferencias filogenéticas pronunciadas entre algunas ramas de un árbol genético y sus linajes principales se encuentran codistribuidos sobre un área amplia.
Categoría III	Árbol genético con divergencia reciente, linajes alopátricos. La mayoría de los haplotipos están cercanamente relacionados y están localizados geográficamente cada uno por separado.
Categoría IV	Árbol genético con divergencia reciente, linajes simpátricos. Este patrón se espera en especies con alto flujo génico cuyas poblaciones no han sido divididas por barreras biogeográficas antiguas.
Categoría V	Árbol genético con divergencia reciente, distribución de linajes variada. Es intermedio entre las Categorías III y IV, y consta de linajes comunes con amplia distribución y de linajes con distribución restringida (cada uno confinado a uno o más linajes locales cercanos).

árboles de genes nucleares para explicar la historia evolutiva de una especie, mientras que en el caso del ADN mitocondrial bastaría un solo árbol (Moore, 1995, 1997; Hoelzer, 1997).

Ciertamente, un éxito de los análisis filogeográficos basados en ADN mitocondrial ha sido la generación de descripciones cada vez mejores de las distribuciones geográficas, relaciones filogenéticas y distancias genéticas entre linajes evolutivos de animales, llevando como consecuencia a un mejor entendimiento de la biogeografía regional y áreas de endemismo (Bermingham y Moritz, 1998). Avise *et al.* (1987) observaron que los factores biogeográficos históricos, los factores ecológicos y la conducta de los organismos han jugado un papel fundamental al moldear la estructura genética de varias especies extintas; a la síntesis de esos factores le llamaron patrones filogeográficos. Estos patrones se agrupan en cinco categorías (tabla 1) y cada una de ellas refleja los diferentes aspectos espaciales y de escala de la estructura genealógica de las poblaciones (Avise, 2000).

Métodos y análisis de la filogeografía

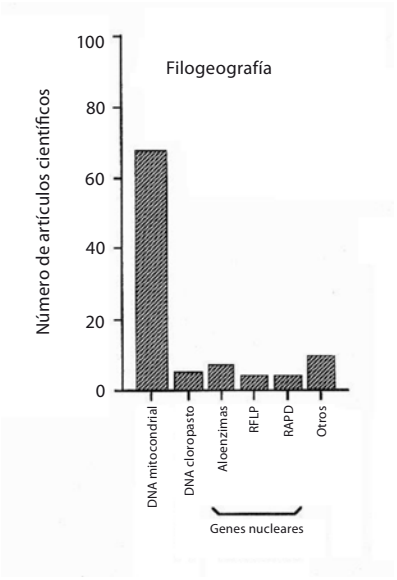
En los estudios filogeográficos se requieren programas computacionales con el fin de obtener resultados robustos sobre la relación entre la estructura poblacional y la distribución geográfica de la especie y su historia. A continuación se explican brevemente algunos de los análisis y programas que se utilizan más comúnmente.

Construcción de árboles

La construcción de árboles para entender los patrones filogeográficos es una parte fundamental, ya que éstos servirán de base para la interpretación y la realización de los análisis siguientes. Usualmente se construyen dos tipos de árboles, unos que son de distancia –neighbor joining (Saitou y Nei, 1987), minimum evolution (Nei y Kumar, 2000), UPGMA (Sneath y Sokal, 1973)- y otros son de parsimonia (veáse el capítulo 4 de este libro). Éstos últimos servirán de base para llevar a cabo los análisis de Neigel (en Avise, 2000) o los análisis de clados anidados (propuesto por Templeton *et al.,* 1995; Templeton 1998; véanse los capítulos 4 y 11 de este libro), que se verán más adelante. También están los árboles probabilísticos, como los de máxima verosimilitud (Felsenstein, 1981) y los obtenidos por inferencia bayesiana (este último es un nuevo método que rápidamente ha tenido gran aceptación por parte de los sistemáticos).

La mayoría de los estudios de filogeografía se basa en datos de secuencias obtenidas de diferentes genes, principalmente mitocondriales (figura 2). Estas secuencias son obtenidas por diferentes métodos en el laboratorio, se alinean con la ayuda de diferentes programas computacionales (por ejemplo ClustalX 1.81 de Thompson *et al.,* 1997; véase también el capítulo 4 de este libro), y se procede a la construcción de árboles.

Figura 2. Porcentaje de artículos de filogeografía que utilizan diferentes marcadores moleculares. Se aprecia que los más utilizados son los mitocondriales (tomado de Avise, 2000)



Para los árboles de distancia, en los programas MEGA 2.1 (Kumar *et al.,* 2001) y PAUP* ver. 4 (Swofford, 1999) es posible obtener matrices de distancia entre las secuencias con diferentes modelos de sustitución de nucleótidos -Jukes y Cantor (Nei y Kumar 2000), Kimura 2 parámetros (Kimura 1980), Tamura Nei (Tamura y Nei 1993), por mencionar algunos- y así determinar el número de haplotipos obtenidos, así como la composición nucleotídica del gen.

Los árboles de parsimonia se construyen con base en los haplotipos obtenidos con varios programas computacionales -Hennig 86 (Farris, 1988), PAUP* ver. 4 (Swofford, 1999), NONA (Goloboff, 1993), PHYLYP (Felsenstein, 1993)- siendo el PAUP* el más utilizado. Cabe mencionar que también dentro

de este programa se pueden obtener algunos árboles de distancia (neighbor joining y UPGMA).

En este capítulo nos referiremos solamente a los árboles de parsimonia obtenidos por medio de PAUP (*Phylogenetic Analysis Using Parsimony*), que utiliza matrices en formato Nexus. Es necesario aclarar que no todos los caracteres son filogenéticamente informativos, por lo que cuando se realice cualquier tipo de análisis se deben involucrar solamente este tipo de caracteres (véase el capítulo 4 de este libro).

En PAUP existen tres opciones para la búsqueda del árbol más parsimonioso, que son la búsqueda heurística, la de *branch-and-bound* y la exhaustiva. La mejor búsqueda será aquella que es exhaustiva, pero este tipo de análisis sólo se puede llevar a cabo cuando se trabaja con menos de 10 taxones o haplotipos. Cuando el número es mayor, puede llevar mucho tiempo y las computadoras pueden carecer de memoria suficiente para realizarla, ya que se calcula que con más de 10 taxones el número de árboles enraizados más parsimoniosos puede rebasar los 34 millones. Por lo anterior, se aconseja realizar tanto la búsqueda heurística como la de *branch-and-bound*. Generalmente la topología encontrada en ambos árboles es muy parecida. Una búsqueda heurística es muy rápida y es conveniente cuando el número de taxones y caracteres es grande, como es el caso de los caracteres moleculares. Existen varios algoritmos de búsqueda heurística, uno comúnmente usado es el *stepwise addition*, muy parecido al *branch-and-bound*. Generalmente es necesario llevar a cabo la prueba estadística del "bootstrap" para poner a prueba el apoyo que tienen las ramas del árbol o los árboles encontrados y se consideran con buen apoyo aquellas que tienen más del 50% (Hillis y Bull, 1993). Cuando se encuentran varios árboles lo que se sugiere es obtener un árbol de consenso (véase el capítulo 4 de este libro).

Genética de poblaciones

Las variaciones moleculares genéticas (alélicas o nucleotídicas) dentro de las secuencias, proporcionan la descripción de la variación genética de los organismos (Hedrick, 2000). Una vez que se ha obtenido el árbol filogenético de los haplotipos de las diferentes poblaciones, podemos utilizar esta variación a la luz de la genética de poblaciones para conocer las relaciones evolutivas entre los alelos o haplotipos, así como las frecuencias de los alelos y sus distribuciones espaciales (Templeton, 1998).

La genética de poblaciones se encarga de estudiar cómo se aplican las leyes de Mendel y otros principios genéticos a las poblaciones naturales. Es

esencial para el entendimiento de la evolución, ya que ésta es el resultado de los cambios en la composición génica de las poblaciones. Debido a esto, los genetistas de poblaciones buscan el entendimiento y predicción de los efectos génicos (segregación, recombinación, transposición y mutación) en una población (Hartl y Clark, 1989). Sin embargo, las poblaciones naturales son dinámicas en muchos sentidos: con el tiempo cambian en tamaño, locación y densidad; en el espacio pueden fragmentarse y/o unirse con otras. Para poder estudiar la estructura de las poblaciones naturales se han tomado dos direcciones: la descripción gráfica de la ramificación evolutiva de los linajes de las secuencias de ADN y los modelos matemáticos de la estructura poblacional (Hey y Machado, 2003).

Dentro de los modelos matemáticos, los estadísticos de *F* introducidos por Wright (1951) describen la estructura génica de organismos diploides en términos de tres estadísticos *F* o correlaciones alélicas, F_{is}, F_{it}, F_{st}, los cuales se describen como sigue:

F_{is}: Correlación entre alelos homólogos en los individuos con referencia a la población local.

F_{it}: Correlación alélica correspondiente con referencia a la población local. Éstos dos han sido llamados *índices de fijación* y sus valores pueden interpretarse como la desviación de las frecuentas genotípicas esperadas en Hardy-Weinberg dentro de las poblaciones locales y la población total (Avise, 1994).

F_{st}: se interpreta como la varianza de la frecuencia alélica entre poblaciones, estandarizada al valor máximo posible dada la media observada de frecuencias alélicas.

Estos estadísticos *F* (Wright, 1931) son una herramienta que utiliza las frecuencias alélicas obtenidas de diferentes localidades geográficas para estudiar el flujo génico o para cuantificar la subdivisión de la población, ya que muestra una relación no lineal entre la cantidad de flujo génico y el grado de diferenciación genética entre las subpoblaciones. Dentro de los análisis es muy utilizado el estadístico F_{st} para obtener el índice de diferenciación global y entre las poblaciones de una misma especie (véase el capítulo 2 de este libro).

Cuando una población llega a un estado de equilibrio como resultado del balance entre la mutación, la deriva génica y el intercambio genético, el flujo génico dentro y entre las poblaciones reales puede expresarse como una tasa de migración *m*. Ésta se define como la proporción de alelos en cada

generación que es de origen migratorio. Aún cuando es difícil de separar la influencia de la deriva génica y del flujo génico, el producto de *Nm* (donde *N* es el tamaño de una población subdividida equitativamente bajo un "modelo de isla" y que intercambia alelos con igual probabilidad) puede estimarse a partir de los estadísticos *F* de Wright, de los alelos privados (aquellos que se encuentran solo en una población) y de las filogenias alélicas (Avise, 1994; Hey y Machado, 2003; ver capítulo 2).

Estos estadísticos se obtienen en diferentes programas computacionales, como el programa Arlequin ver. 2000 (Schneider *et al.,* 2000) y el DnaSP 2.52 (Rozas y Rozas, 1997).

Frecuencias de distribución de diferencias pareadas de secuencias

Cuando el tamaño de muestra lo permite, se pueden inferir aspectos de la historia reciente de las poblaciones al llevar a cabo análisis basados en la teoría de coalescencia (Slatkin y Hudson, 1991), ya que ésta proporciona métodos para detectar eventos en el pasado de las poblaciones, tales como el incremento exponencial en tamaño (Zink y Blackwell-Rago, 2000). Mientras mayor sea el tamaño poblacional, el tiempo de coalescencia entre dos secuencias de ADN (tiempo en que ambas convergen en la secuencia que les dio origen) es más grande. En cambio, en poblaciones pequeñas, el tiempo de coalescencia es menor, ya que pueden estar más cercanamente relacionadas (Rogers, 2002; véase también el capítulo 4 de este libro).

Con base en las distancias génicas entre los individuos y suponiendo cierta tasa evolutiva, se pueden obtener estimados provisionales de tiempo absoluto de coancestría materna (cuando se trabaja con ADN mitocondrial). Estos estimados de tiempo acumulados para muchos individuos dan como resultado frecuencias pareadas de distribuciones de tiempos de coancestría, también llamadas distribuciones *mismatch* que se presentan como histogramas en donde se muestra el número de diferencias de sitios nucleotídicos (o restricción) observados entre pares de haplotipos. Es usualmente multimodal (figura 4) en muestras tomadas de una población en equilibrio demográfico, pero es generalmente unimodal en poblaciones que han sufrido una reciente expansión demográfica (Schneider *et al*., 2000).

De manera más sencilla, la distribución *mismatch* cuenta el número de diferencias entre cada par de secuencias en una población (Rogers, 2002) y el resultado permite construir un histograma. Los datos de este histograma se

comparan con el modelo de una población en expansión repentina. Con una prueba de χ-cuadrada estadísticamente significativa se acepta que la diferencia entre ambas no es debida al azar, y se observa en la estructura de la gráfica si las poblaciones experimentaron recientemente un crecimiento exponencial (Zink y Blackwell-Rago, 2000).

Por ejemplo, Zink *et al.,* (2001) realizaron un estudio con genes mitocondriales ND2 y región control I y II para documentar los patrones filogeográficos de seis especies distribuidas en las zonas áridas de Norteamérica y valorar la consistencia entre ellas. Como parte de sus resultados obtuvieron gráficas de *mismatch* para cada especie. En la figura 3 se observa la gráfica para *Polioptila melanura* y se aprecia que la curva es unimodal, por lo que corresponde a un modelo de reciente expansión (Slatkin, 1991; Rogers y Harpending, 1992). Por otra parte, en el trabajo de Zink y Blackwell-Rago (2000) sobre los límites de especie del cuitlacoche piquicurvo, *Toxostoma curvirostre*, observaron los dos tipos de distribución, unimodal y multimodal. Las poblaciones del este de su área de distribución, al presentar una distribución unimodal, probablemente están en expansión demográfica mientras que la población del oeste probablemente pasa por un periodo de estabilidad en el tamaño de su población (figura 4).

Una ventaja de esta aproximación es que en algunos casos un árbol de genes puede carecer de información acerca de la estructura del clado, porque muchos nodos se encuentran temporalmente cerrados, sin embargo, un

Figura 3. Distribución *mismach* de *Polioptila melanura* (tomado de Zink *et al.*, 2001)

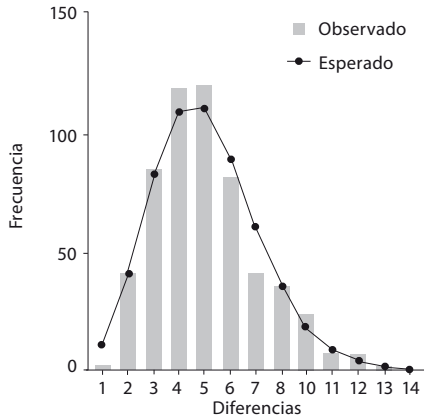
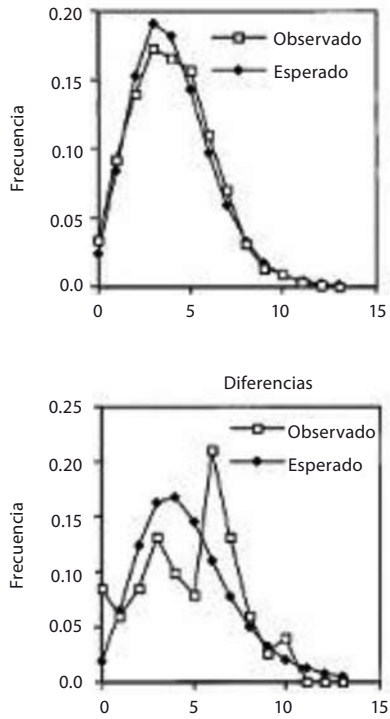


Figura 4. Distribución *mismatch* unimodal y bimodal de *Toxostoma curvirostre* en una población con expansion rápida (a) y la otra en equilibrio (b) (Tomado de Zink y Blacwell-Rago, 2000)



histograma de frecuencia de los tiempos de coancestría puede proyectar la historia demográfica de la población (Avise, 2000). Las diferencias pareadas de las secuencias y la prueba de χ-cuadrada se pueden calcular en el programa Arlequin ver. 2000 (Schneider *et al.,* 2000).

Análisis de clados anidados

En los estudios filogeográficos se han estudiado los efectos de la historia y de la estructura poblacional mediante los análisis de autocorrelación espacial y de componentes principales, sin embargo, como se basan en distancias génicas son análogos a las aproximaciones fenéticas.

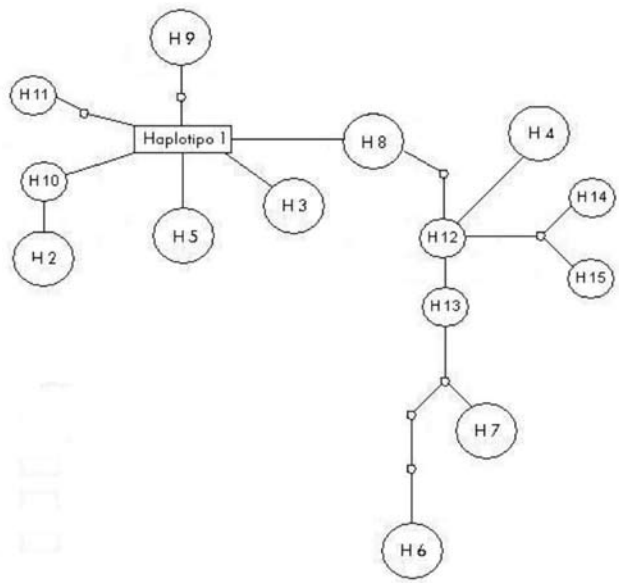
El análisis de clados anidados es una aproximación propuesta por Alan Templeton y sus colegas (1987), quienes se basaron en análisis de enzimas de restricción, midiendo la actividad de la NADH (alcohol deshidrogenasa) de *Drosophila melanogaster* y que después de una serie de artículos fue establecido como método en el estudio del ajolote *Ambystoma tigrinum* (Templeton, 1995).

El análisis de clados anidados (NCA) fue diseñado para distinguir entre los distintos procesos históricos que pudieron influenciar la distribución geográfica actual de la variación génica como se observa en una topología (Hey y Machado, 2003). Para obtener una hipótesis se utiliza un árbol de haplotipos como base para definir una serie jerárquica de clados anidados (ramas dentro de ramas) por medio de la utilización de una serie explícita de reglas de anidación (Templeton *et al.,* 1987; Templeton *et al.,* 1995; Templeton, 2004).

Para anidar, es necesario utilizar el algoritmo elaborado por Templeton *et al.* (1992) que estima la red de haplotipos (figura 5), éste se encuentra en el programa TCS (Clement *et al.,* 2000). Para utilizar este programa no es necesario colapsar las secuencias en haplotipos (todos los individuos que presenten la misma secuencia) ya que esto lo hará el programa así como la estimación de frecuencias para cada haplotipo. En el programa TCS se obtendrá como resultado una red (o árbol) de haplotipos y una gráfica en donde se muestra el número de mutaciones que ocurre entre ellos. Este cladograma de haplotipos se convierte en un diseño anidado (figura 6) en el cual los haplotipos ("clados de 0-pasos") separados por una mutación se agrupan para formar un "clado de un paso" desde los extremos hasta el interior del árbol; posteriormente estos clados de un paso que estén separados a su vez por una mutación se agrupan en un "clado de dos pasos" y así sucesivamente, hasta que el siguiente nivel de anidación abarque todo el árbol (Templeton 1995). Los clados anidados resultantes se denominan como "C-N" en donde "C" es el nivel de anidación del clado y "N" es el número de un clado en particular en un nivel de anidación dado. Esta serie de ramas anidadas constituye un diseño estadístico evolutivo que originalmente fue diseñado para investigar las relaciones entre genotipo y fenotipo (Templeton, 1987).

Una vez que el clado anidado ha sido construido mediante la topología de la red de haplotipos, la prueba más simple para asociación geográfica trata a cada localidad muestreada como una variable categórica. En el programa GeoDis (Posada *et al.* 2000) se implementa el análisis cladístico anidado, ya que al introducir las coordenadas geográficas de las poblaciones se calculan dos distancias principales: distancia de clado (D_c) y distancia de nido (D_n). La primera mide la

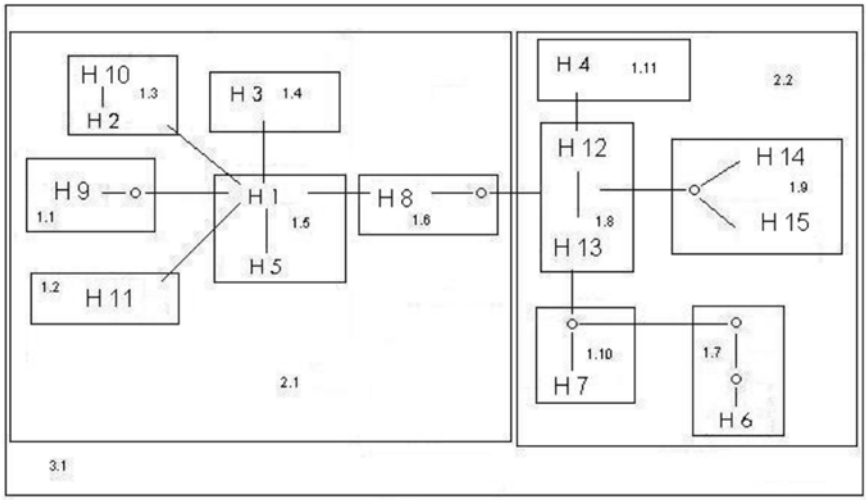
Figura 5. Red de haplotipos elaborada en el programa TCS. Cada haplotipo se relaciona con otro por un solo cambio en la secuencia de ADN; los círculos pequeños muestran haplotipos inexistentes que pudieron no haber sido muestreados o que existieron en algún momento y que desaparecieron



dispersión geográfica de cada uno de los clados mientras que la segunda mide como está distribuido el clado geográficamente con relación a otros clados en el mismo nivel de categoría anidada. Dentro del análisis, también se calculan las diferencias que existen entre el clado y la distancia del clado anidado para los clados exteriores (*tip*) contra los clados interiores (*interior*). Los clados exteriores son aquellos clados conectados al resto del cladograma por solo una mutación, mientras que los interiores son aquellos conectados por dos o más mutaciones. El contraste entre interior-exterior corresponde a un contraste entre joven y viejo, y entre poco distribuido o raro y común. El archivo resultante de este análisis consiste en los estadísticos calculados y sus valores asociados permutacionales de *P*, los cuales mediante las tablas de inferencia de Templeton (2004) pueden interpretar estos resultados en términos biológicos.

Las tablas de inferencia están basadas en tres factores biológicos que Templeton *et al*. (1995) consideran como causantes de asociaciones espa-

Figura 6. Red de haplotipos de la figura anterior, dividida en clados de anidación de diferentes niveles. Los cuadros más pequeños (numerados del 1.1 al 1.11) son clados de primer nivel o de un paso, éstos se encuentran agrupados en cuadros más grandes (numerados con el 2.1 y 2.2) que representan clados de segundo nivel; estos dos clados se agrupan en un cuadro mayor (3.1) que constituye un clado de tercer nivel



cio-temporales significativas que provocan la variación de los haplotipos. El primer factor es el flujo génico restringido, particularmente debido a aislamiento por distancia (véase el capítulo 2 de este libro), el segundo factor son los eventos de fragmentación típicamente alopátricos y el tercer factor es la expansión de rango, incluida la colonización, ya que cuando ocurre, todos los haplotipos que se encuentran en la población ancestral se expandirán geográficamente. La utilización de las tablas de Templeton ha sido debatida debido a que son difíciles de leer y subjetivas (Knowles y Maddison, 2002), sin embargo Templeton (2004) afirma que esta clave representa una separación formal de la prueba estadística contra la interpretación biológica que debe estar acompañada por otras pruebas como verosimilitud o la utilización de genes múltiples.

Estudios mexicanos

Las aves han sido y son actualmente uno de los grupos que más se utilizan para los estudios filogeográficos. Primero, con las aloenzimas, se tenía evidencia de

que no hay o existe poca estructura genética en sus poblaciones en comparación con la estructura genética de especies de otras poblaciones de vertebrados en la misma zona (Zink, 1997). Sin embargo, recientemente las poblaciones de varias especies de aves se han revalorado desde una perspectiva filogenética y se ha descubierto gran variedad de patrones de ADN mitocondrial, aunque no todos serán mencionados aquí (ver Avise, 2000). Para el caso de las aves de México, los estudios de este tipo han sido muy pocos y se han limitado a las poblaciones residentes fuera del país (Moore *et al.,* 1991; Fletcher y Moore, 1992; Johnson y Marten, 1992; Johnson, 1995; Milot *et al.,* 2000; Pruet *et al.,* 2000; Bulgin *et al.,* 2003); no es sino recientemente que algunos grupos de trabajo (*e.g.* Museo de Zoología, Facultad de Ciencias, UNAM) han tratado de abordar el tema.

Dentro de los estudios realizados en especies de aves residentes a México se han encontrado varios patrones filogeográficos que se encuentran en algunas de las categorías propuestas por Avise (2000; tabla 1). Así, podemos encontrar especies cuya estructura filogeográfica corresponde con su distribución geográfica y su fenotipo, hasta especies cuyo fenotipo y estructura poblacional no se relacionan entre sí ni con la distribución geográfica de la especie. En cuanto a las especies migratorias, los estudios filogeográficos son pocos y la mayoría se encuentran orientados al área de reproducción del ave (Estados Unidos de América y Canadá) sin considerar la zona de invierno.

En los estudios filogeográficos realizados con aves residentes y que no presentan ninguna estructura filogeográfica, por lo que corresponden a la Categoría III ó IV, se encuentran las de distribución disyunta como *Melospiza melodia* (Zink y Dittmann, 1993) y *Petrochelidon fulva* (Kirchman *et al.,* 2000). La especie *Melospiza melodia* tiene una amplia distribución desde las islas Aleutianas y la costa este de Estados Unidos de América hasta el centro de México y ha sido dividida en numerosas subespecies. Sin embargo, el estudio realizado con ADN mitocondrial no revela ninguna estructura genética para la especie (a diferencia de la morfología) ni ningún orden en la distribución geográfica de los alelos a lo largo de toda su distribución, incluidas las localidades mexicanas. Esto refleja un reciente establecimiento de la población después de una expansión de toda su área de distribución.

En el estudio realizado en *Petrochelidon fulva*, que tiene distribución disyunta en el centro-sur de Estados Unidos de América, Florida, la península de Yucatán, las islas del Caribe, Ecuador y Perú, se encontró que para México (centro-norte del país y península de Yucatán), la diferenciación es mínima y no se podría hablar de unidades separadas (subespecies *P. f. pallida*

y *P. f. citata*, respectivamente), ya que mantienen un flujo génico constante y realizan migraciones anuales.

En los estudios donde la estructura filogeográfica tiene una correspondencia parcial con la distribución geográfica, Categoría I ó III, se encuentran el pavo *Meleagris gallipavo* (Mock *et al.,* 2002) y el chipe *Wilsonia pusilla* (Kimura *et al.,* 2002). Para México se estudiaron tres poblaciones norteñas de *Meleagris gallopavo* (*M. g. mexicana, M. g. merriami* y *M. g. intermedia*) y se encontró que se existen suficientes diferencias a nivel genético como para hacer la separación entre subespecies; esta separación puede deberse a la presencia de las cadenas montañosas mexicanas y el desierto de Sonora y Chihuahua como barreras biogeográficas. Sin embargo, los resultados deben tomarse con ciertas reservas, ya que a pesar de que se tuvo mucho cuidado de no incluir ejemplares reintroducidos, no se incluyó el resto de la distribución de la especie en el sureste del país, que corresponde a la subespecie *M. g. gallopavo*.

Para la especie migratoria *Wilsonia pusilla*, se realizó un estudio filogeográfico en poblaciones reproductivas y de invierno. Las poblaciones reproductivas se distribuyen en casi todo el territorio de Estados Unidos de América, y las de invierto abarcan desde la mitad del altiplano central de México, hacia el sur hasta Centroamérica. En las poblaciones reproductivas en encontraron haplotipos que dividen a la especie en este y oeste de EUA. Para la distribución invernal en México se observó que la mayoría de las aves que llegan son del haplotipo del oeste, y únicamente tres localidades muestreadas del sureste del país presentan haplotipos del este del área de reproducción. Por lo tanto, se tiene una estructura filogeográfica muy marcada en las áreas de reproducción, que se conservan en su mayoría en el área de invierno, y que al parecer no se encuentran influenciadas por la geografía.

También se encontró que *Toxostoma lecontei* (Zink y Blackwell-Rago, 1997) tiene una estructura filogeográfica congruente con la genética y la distribución, por lo que pertenece a la Categoría I ó III. Esta especie se distribuye en las zonas desérticas del suroeste de Estados Unidos de América y noroeste de México incluyendo Baja California Norte. Se realizó un estudio filogeográfico con ADN mitocondrial (gen citocromo b) de muestras provenientes de nueve localidades de su área de distribución y se obtuvo una separación geográfica en la zona del desierto del Vizcaíno en Baja California Norte. Esta zona corresponde a la subespecie *T. l. arenicola* que se separa de *T. l. lecontei* que ocupa el resto del área de distribución; se distinguen como unidades diferentes y los autores las proponen como dos especies distintas, tomando como base también las diferencias morfológicas.

Dentro de los estudios de filogeografía comparativa, se han realizado dos estudios con aves mexicanas que extienden sus distribuciones hasta Estados Unidos de América (Zink, 1996) y las tierras áridas de Norteamérica (Zink *et al.,* 2001). En estos estudios utilizan las aves como modelo para explicar los patrones filogeográficos con base en la variación del ADN mitocondrial y con ello evaluar el papel que tuvieron diferentes fuerzas evolutivas (flujo génico, cuellos de botella, barreras históricas y ecológicas) en los patrones espaciales de las especies. Los estudios filogeográficos comparativos son interesantes porque si se encuentra un patrón filogeográfico general de diferentes especies de aves, podría suponerse que tuvieron una misma historia geográfica, e incluso al utilizar una gran diversidad de grupos (por ejemplo aves terrestres, invertebrados marinos, peces de agua dulce y peces de agua salada; Avise, 1998), encontrar patrones filogeográficos generales hace suponer, sin lugar a dudas, una historia geográfica común.

Zink (1996) trabajó con cinco especies de aves en México y trató de describir y comparar los patrones filogeográficos que presentaban con base en su variación del ADN mitocondrial. Entre ellas están representadas cuatro especies de aves paserinas y una no paserina (ganso), con una distribución amplia y simpátrica en las regiones de Norteamérica. Encontró que dos de las especies estudiadas, el ganso (*Branta canadensis*) y el gorrión (*Passerina iliaca*) mostraron patrones filogeográficos incongruentes entre sí, mientras que las tres especies restantes (*Melospiza melodia, Spizella passerina* y *Agelaius phoeniceus*) carecen de un patrón. Estos resultados sugieren que la avifauna de Norteamérica está compuesta por especies que tuvieron diferentes historias en el pasado, es decir, han existido periodos de expansión reciente (para las tres especies que no muestran ningún patrón filogeográfico) así como importantes eventos de aislamiento que produjeron diferentes patrones filogeográficos en las otras dos especies, aunque incongruentes entre sí.

Zink *et al.* (2001) trataron de encontrar patrones filogeográficos en especies con distribuciones que abarcan el suroeste de Estados Unidos de América, Baja California y el desierto de Sonora-Chihuahua en México, definiendo en algunos casos patrones filogeográficos similares y en otros no. Utilizaron secuencias de ADN mitocondrial (región control y ND2) y encontraron que para *Campylorhynchus brunneicapillus* y *Auriparus flaviceps* existe una interrupción filogeográfica cerca de los 30ºN, lo que coincide con lo que habían previamente reportado para *Toxostoma lecontei* (Zink y Blackwell-Rago, 1997) y que puede explicarse por la existencia de un brazo marino que existió en Baja California hace un millón de años que provocó un aislamiento entre el

norte y el sur. En contraste con lo anterior, *Polioptila californica* carece de tal interrupción filogeográfica y la expansión de su área de distribución hacia el norte ha sido reciente. También *T. curvirostre* y *Pipilo fuscus* muestran divisiones genéticas entre los desiertos de Sonora y Chihuahua, mientras que *C. brunneicapillus*, *P. melanura* y *A. flaviceps* no lo hacen; para estas últimas una explicación probable es que la colonización de su actual área de distribución sea reciente y no refleje la distribución que tuvieron en el pasado. Por lo anterior, concluyen que no todas las especies de Baja California han tenido la misma historia, ni han seguido los mismos caminos, por lo que no se puede, aún, hacer una generalización de la filogeografía de Baja California, pero este tipo de estudios (filogeografía comparada) pueden proveer de información para determinar si algunas especies pudieron haber coexistido en algún momento de su historia geográfica.

El hecho de que se hayan encontrado patrones filogeográficos muy diferentes en las especies de aves mexicanas aquí revisadas nos sugiere que la historia biogeográfica de México para las aves es muy compleja y que no está regida por eventos únicos, sino por una muy compleja interacción de fenómenos y procesos cuyo resultado es la diversidad de aves que tenemos en el país y la reconstrucción de diferentes escenarios filogeográficos.

Actualmente, en el laboratorio de Sistemática Molecular del Museo de Zoología de la Facultad de Ciencias de la UNAM se realizan varios proyectos sobre límites de especies y filogeografía de aves mexicanas como un gran esfuerzo y contribución al conocimiento de las aves de México. Los estudios que se llevan a cabo incluyen gran variedad de órdenes y familias, particularmente de complejos de especies como el colibrí piquiancho *Cynanthus latirostris* (García-Deras, 2003), el colibrí garganta amatista *Lampornis amethystinus* (Cortés-Rodríguez, 2003) y el carpintero arlequín *Melanerpes formicivorus* (Honey-Escandón, 2002), entre otros.

De los resultados de estos trabajos se establece que la descripción de las subespecies ha sido subjetiva, sin análisis estadísticos morfológicos rigurosos que las apoyen (*e.g.* Montes de Oca, 1875; Ridgway, 1911; Moore, 1939; Berlioz, 1937; Binford, 1985, 1989). En el caso de *Cynanthus latirostris* se estudiaron cinco de las seis subespecies descritas y sólo dos, *latirostris* y *doubledayi*, son unidades diagnósticas reconocibles tanto en lo morfológico como en lo molecular (García-Deras, 2003). Se obtuvo una distribución *mismatch* multimodal para cada taxón lo que sugiere que las muestras provienen de poblaciones en equilibrio demográfico después de haber pasado probablemente por un cuello de botella (Schneider *et al.*, 2000; Rogers, 2002). El valor de F_{st} obtenido para

el género *Cynanthus* fue 0.04495 lo que nos sugiere que existe flujo génico y que las poblaciones están moderadamente diferenciadas (Zink y Blackwell-Rago, 2000; Baker *et al.*, 2003) mientras que la diversidad nucleotídica π promedio fue igual a 0.2094. El valor de F_{st} obtenido para el género *Cynanthus*, la reconstrucción filogenética de los haplotipos y la diferencia en el color del plumaje del macho en *Cynanthus doubledayi* sugieren que se trate como una especie diferente de *C. latirostris* por lo que en este trabajo se reconocieron 3 especies del género *Cynanthus*: *C. sordidus*, *C. latirostris* y *C. doubledayi*, una más de las previamente reconocidas. Sin embargo, se hace la aclaración de que de haber incluido la sexta subespecie endémica de las Islas Tres Marías en el análisis, conformaría otra unidad diagnóstica reconocible. De esta manera, el género *Cynanthus* estaría conformado por cuatro especies en lugar de las dos reconocidas actualmente.

El colibrí garganta amatista (*Lampornis amethystinus*) es residente de los bosques de pino-encino y mesófilos de montaña (900-3400 msnm) desde Nayarit y sur de Tamaulipas, a lo largo de las tierras altas de México y Guatemala, hasta Honduras (AOU, 1998). Se han descrito siete subespecies a lo largo de toda su distribución y se realizó un estudio filogeográfico para conocer si estas

Figura 7. Distribución *mismatch* bimodal de *Cynanthus sordidus* que sugiere que la población está en un equilibrio demográfico (García-Deras, 2003)

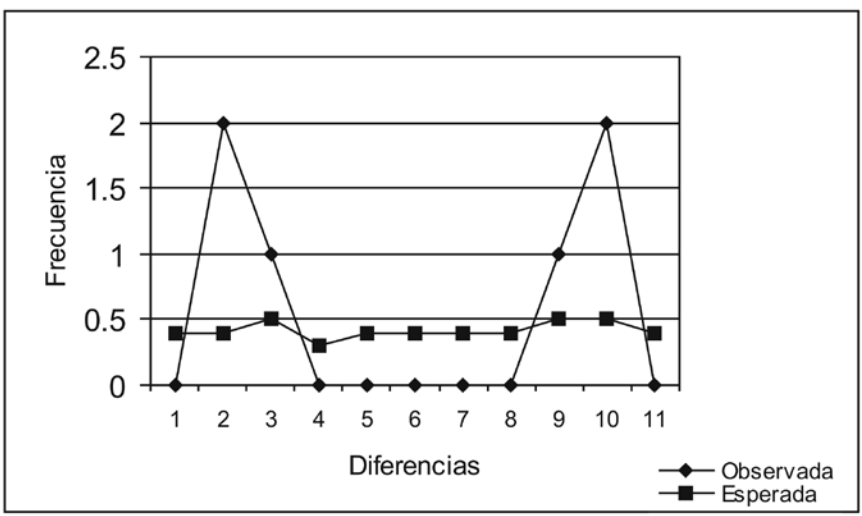


Figura 8. Distribución *mismatch* multimodal de *Cynanthus doubledayi* que sugiere que la población está en un equilibrio demográfico (García-Deras, 2003)

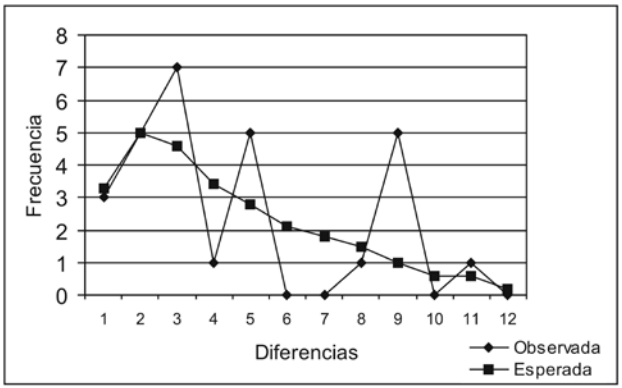
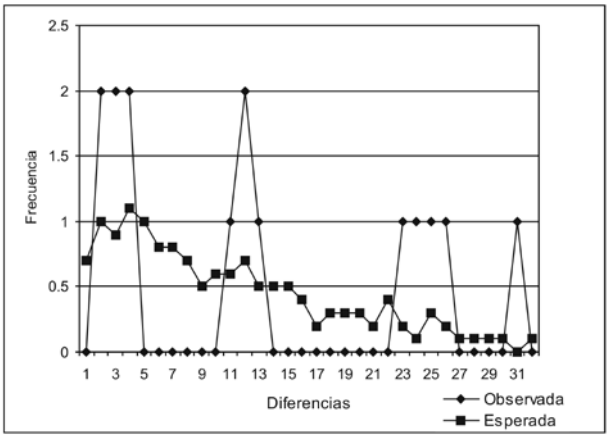


Figura 9. Distribución *mismatch* multimodal de *Cynanthus latirostris* que sugiere que la población está en un equilibrio demográfico (García-Deras 2003)



subespecies son en realidad genéticamente distintas, utilizando 73 individuos pertenecientes a 27 localidades, y donde se encontraron 16 haplotipos. De estos haplotipos 6 pertenecen a la subespecie *Lampornis amethystinus salvini*

distribuida desde Chiapas hasta Honduras, morfológicamente muy similar a *L. amethystinus amethystinus* al presentar una tonalidad de la garganta rosa oscuro a diferencia de otra subespecie *Lampornis amethystinus margaritae* (de Oaxaca y Guerrero) que presenta la garganta morada. Por medio de un análisis de clados anidados (figura 10) se observó que la diferenciación genética entre *amethystinus* y *salvini* fue causada por una barrera geográfica, el Istmo de Tehuantepec, capaz de moldear la variabilidad genética de esta especie. El patrón filogeográfico que se encontró en este estudio pertenece a la categoría filogeográfica IV Avise (2000); los valores de diferenciación genética obtenidos para las poblaciones del centro de México así como para las de Chiapas-Centroamérica presentan una diferenciación genética moderada, lo que indica que presenta linajes simpátricos con conexiones evolutivas recientes (Cortés-Rodríguez, 2003).

El carpintero arlequín, *Melanerpes formicivorus,* se distribuye en tierras altas con vegetación de pino-encino, desde el oeste y centro-sur de Estados Unidos de América hacia el sur en las cadenas montañosas de México hasta las sierras de Centroamérica. Tiene algunas poblaciones aisladas en el sur de Baja California Sur, el centro de Costa Rica y los Andes de Colombia. Se realizó un análisis filogeográfico para esta especie con muestras de algunos estados

Figura 10. Clados anidados para los 16 haplotipos de ADN mitocondrial identificados en *L. amethystinus*, donde los ceros indican haplotipos intermedios que no se encontraron en el análisis. Cada línea representa un cambio mutacional sencillo que interconecta dos haplotipos con una probabilidad mayor de 95%. Los clados de segundo nivel indican aislamiento por una separación geográfica

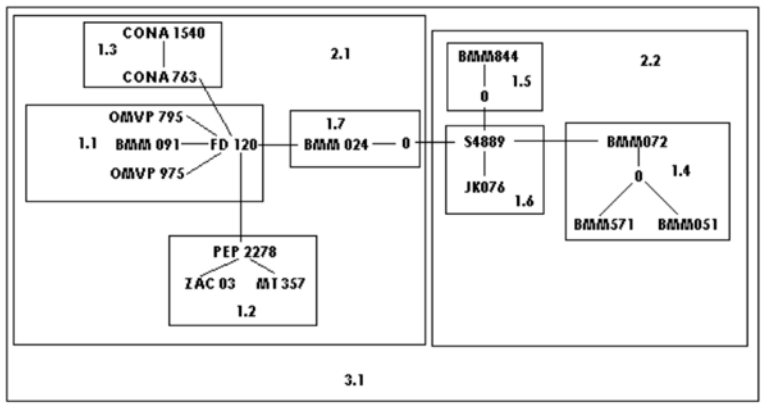
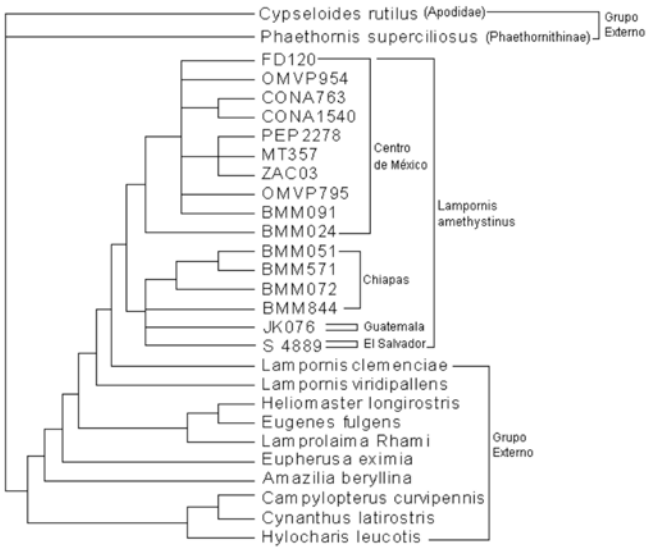


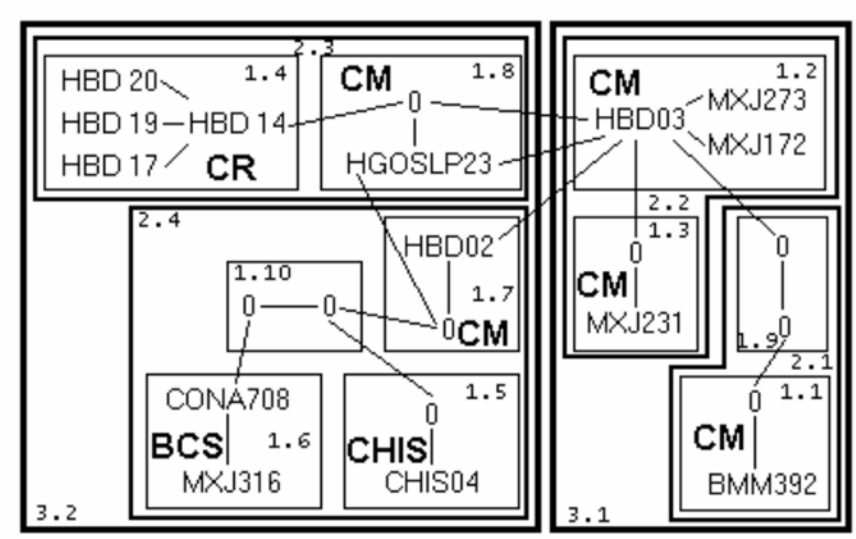
Figura 11. Árbol de parsimonia construido con el programa PAUP para 16 haplotipos del gen mitocondrial ND2 del colibrí *Lampornis amethystinus*



de México y Costa Rica, y se encontraron cuatro unidades de historia: *M. f. formicivorus-bairdi* (centro de México y Estados Unidos de América), *M. f. striatipectus* (Costa Rica), *M.f. angustifrons* (Baja California Sur) y *M. f. lineatus* (Chiapas). El patrón filogeográfico encontrado en este trabajo corresponde al V donde las unidades son el resultado del aislamiento de las poblaciones, por un lado, lo que condujo a diferentes historias evolutivas de cada una de las zonas más aisladas de la especie (Baja California Sur, Costa Rica y Chiapas). Por otro lado, también posee linajes comunes ampliamente distribuidos, como los haplotipos del centro de México. El estado actual de la especie es una mezcla de varios procesos que actúan de forma diferente entre los grupos de la misma (Honey-Escandón 2002), por lo que no se puede reconstruir una historia única para la especie a partir de la información proporcionada en los clados anidados (figuras 12 y 13).

En conclusión, de acuerdo con nuestros resultados, podemos inferir que tanto para *M. formicivorus* como para *Lampornis amethystinus* el Istmo de Tehuantepec ha sido una barrera geográfica importante que ha conducido a la separación de unidades de historia diagnosticables, provocando una

Figura 12. Árbol de clados anidados realizado con el programa TCS para los 15 haplotipos de una porción del gen mitocondrial ND2 del carpintero *Melanerpes formicivorus.* La anidación de los clados de segundo y tercer nivel no corresponde con el aislamiento de los grupos encontrados en el árbol de parsimonia. En cada clado de primer nivel, las iniciales indican la localidad donde provienen los haplotipos: CR Costa Rica, CM Centro de México, BCS Baja California Sur y CHIS Chiapas

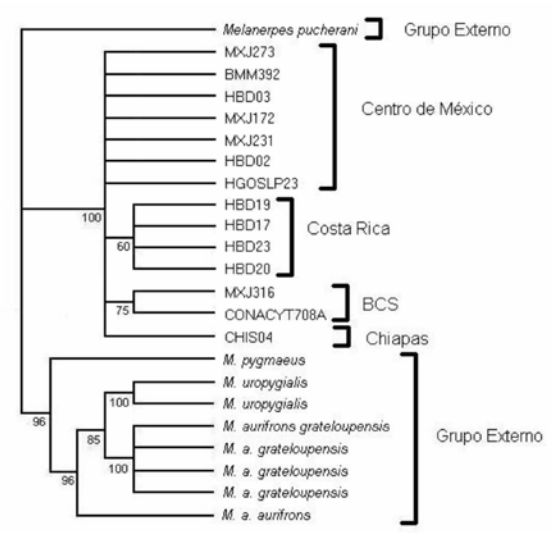


diferenciación morfológica seguida de una genética (y no al revés). Este patrón filogeográfico ya había sido planteado por varios autores con base en la morfología de colibríes (Binford, 1989; Howell y Webb, 1995) y otras especies de aves como *Chlorospingus ophtalmicus*, *Streptoprocne zonaris* y *Turdus infuscatus* (Binford, 1989; Howell y Webb, 1995; Sánchez-González, 2002). En esta ocasión lo documentamos genéticamente, por lo que la hipótesis se hace más robusta.

Otros trabajos en desarrollo tratan con especies de distribución neotropical, como la tucaneta esmeralda, *Aulacorhychus prasinus* y el colibrí fandanguero *Campylopterus curvipennis.*

El colibrí *Campylopterus curvipennis* tiene una distribución disyunta, en la vertiente del Golfo y en la Península de Yucatán hasta el este de Honduras. Se piensa que en estas dos áreas puede haber dos grupos diferentes: *C. c. curvipennis* y *C. c. pampa* (respectivamente) (AOU, 1998). Además, algunos autores sugieren que junto con *C. excellens*, que se distribuye en la Sierra de

Figura 13. Árbol de parsimonia (consenso de mayoría del 50% de 28 árboles) obtenido en el programa PAUP, con búsqueda heurística, de 14 haplotipos de una porción del gen mitocondrial ND2 del carpintero *Melanerpes formicivorus*. Se señalan los principales grupos obtenidos. El grupo Centro de México se refiere a localidades situadas en Jalisco, Michoacán, Durango, Baja California Norte, Hidalgo, Guerrero, Oaxaca, Querétaro y Zacatecas



los Tuxtlas, Veracruz y noroeste de Chiapas, forma una superespecie (Sibley y Alquist, 1990), o que incluso *C. excellens* es otra subespecie de *C. curvipennis*. Por lo tanto, se analizará este complejo para poder encontrar diferencias genéticas entre estos tres grupos y una estructura filogeográfica que proporcione mayor información sobre su historia evolutiva.

En cuanto a la tucaneta esmeralda *Aulacorhynchus prasinus*, su distribución va desde el centro de México hasta Panamá, los Andes de Colombia, oeste de Venezuela, Ecuador y Perú (AOU, 1998). Se encuentra asociada a bosques mesófilos de montaña por lo que presenta una distribución fragmentada, restringida o interrumpida con condiciones ecológicas específicas. Dado que sus poblaciones presentan patrones de variación discreta en tamaño, coloración del plumaje y del pico, algunas de las 16 subespecies reconocidas (Peters, 1948; Winker, 2000) se consideran especies distintas y además existen

controversias sobre el estatus taxonómico de la especie. Por lo tanto, se estudia este complejo para evaluar límites de especies entre las diversas formas a través de diferencias genéticas y morfológicas, así como inferir aspectos de la historia evolutiva y de distribución geográfica de las mismas a partir de análisis filogeográficos.

Gracias a la Colección de Tejidos de Aves del Museo de Zoología, que consta de más de 5000 tejidos catalogados (número que incrementa constantemente), en un futuro cercano puede surgir mayor número de proyectos completamente mexicanos y otros en colaboración con investigadores extranjeros. Estos primeros estudios realizados en México son fruto del esfuerzo de varios años de investigación y desarrollo y representan un primer resultado del trabajo de investigadores mexicanos. Se espera la pronta conclusión de los proyectos aquí presentados y que en un futuro se tengan los patrones filogeográficos de varias especies de aves. También se considera la posibilidad de realizar filogeografía comparativa con los patrones que surjan.

Bibliografía

American Ornithologist´ Union. 1998. *Check-list of North American Birds*. AOU.

Avise, J.C. 1994. *Molecular markers, natural history and evolution*. Chapmann & Hall.

Avise, J.C. 1998. The history and purview of phylogeography: a personal reflection. *Molecular Evolution* 7:371-379.

Avise, J.C. 2000. *Phylogeography. The history and formation of species*. Harvard University Press, Cambridge.

Avise, J.C., J. Arnold, R.M. Ball Jr., E. Bermingham, T. Lamb, J.E. Neigel, C.A. Reeb y N.C. Saunders. 1987. Intraspecific phylogeography: the mitochondrial DNA bridge between population genetics and systematics. *Annual Reviews Ecology and Systematics* 18:489-522.

Baker, J., E. López-Medrano, A. Navarro-Sigüenza, O. Rojas-Soto y K. E. Omland. 2003. Recent speciation in the orchard Oriole group: divergence of *Icterus spurius spurius* and *Icterus spurius fuertesi*. *The Auk* 120(3):848-859.

Berlioz, J. 1937. Note sur une collection des oiseaux du Mexique. *Bulletin Museum d'Histoire Naturelle* 9:170-175.

Bermingham, E. y C. Moritz. 1998. Comparative phylogeography: concepts and applications. *Molecular Evolution* 7:367-369.

Binford, L.C. 1985. Re-evaluation of the "hybrid" hummingbird *Cynanthus sordidus* x *C. latirostris* from Mexico. *The Condor* 87(1):148-150.

Binford, L.C. 1989. A distributional survey of the birds of the mexican state of Oaxaca. *Ornithological Monographs* 43:1-418.

Bulgin, N.L., H.L. Gibbs, P. Vickery y A.J. Baker. 2003. Ancestral polymorphisms in genetic markers obscure detection of evolutionarily distinct populations in the endangered Florida grasshopper sparrow (*Ammodramus savannarum floridanus*). *Molecular Ecology* 12:831-844.

Clement, M., D. Posada y K.A. Crandall. 2000. TCS: a computer program to estimate gene genealogies. *Molecular Ecology* 9:1657-1659.

Cortés Rodríguez, M.N. 2003. Filogeografía de *Lampornis amethystinus* Swainson (Aves: Trochilidae). Tesis de Licenciatura. Facultad de Ciencias, UNAM.

Farris, J.S. 1988. Hennig86 inference. Version 1.5. Publicado por el autor, Port Jefferson, Nueva York.

Felsenstein, J. 1981. Evolutionary trees from DNA sequences: a maximum likelihood approach. *Journal of Molecular Evolution* 17: 368-376.

Felsenstein, J. 1993. PHYLIP version 3.5c. Univ. of Washington, Seattle, Washington.

Fletcher, S.D. y W.S. Moore. 1992. Further analysis of allozyme variation in the northern flicker, in comparison with mitochondrial DNA variation. *The Condor* 94:988-991.

García-Deras, G.M. 2003. Límites de especies dentro del complejo *Cynanthus latirostris* (Aves: Trochilidae). Tesis de Licenciatura. Facultad de Ciencias, UNAM.

Goloboff, P. 1993. NONA version 2, publicado por el autor.

Hartl, D.L., A. G. Clark. 1989. Principles of Population Genetics. Sinnauer Associates. Massachussets. USA.

Hedrick, P. 2000. *Genetics of Populations*. Jones & Barlett Publishers, Boston.

Hey, J y C. A. Machado. 2003. The study of structure populations-New hope for a difficult and divided science. Genetics: 4, 535-543.

Hillis, D. M y J. J. Bull. 1993. An empirical test of bootstrapping as a method for assessing confidence in phylogenetic analysis. *Systematic Biology* 42: 182-192.

Hoelzer, G.A. 1997. Inferring phylogenies from mtDNA variation: mitochondrial-gene trees versus nuclear-gene trees revisited. *Evolution* 51(2):622-626.

Honey Escandón, M. 2002. Filogeografía de las poblaciones del carpintero arlequín *Melanerpes formicivorus* (Aves: Picidae). Tesis de Licenciatura, Facultad de Ciencias, UNAM.

Howell, S.N.G. y S. Webb. 1995. A guide to the birds of Mexico. Oxford University Press.

Johnson, N.K. 1995. Speciation in vireos I. Macrogeographic patterns of allozymic variation in the *Vireo solitarius* complex in the contiguous United States. *The Condor* 97:903-919.

Johnson, N.K. y J.A. Marten. 1992. Macrogeographic patterns of morphometric and genetic variation in the sage sparrow complex. *The Condor* 94:1-19.

Kimura, M. 1980. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *Journal of Molecular Evolution* 16: 111-120.

Kimura, M., S.M. Clegg, I.J. Lovette, K.R. Holder, D.J. Girman, B. Milá, P. Wade y T.B. Smith. 2002. Phylogeographical approaches to assessing demographic conectivity between breeding and overwintering regions in neartic-neotropical warbler (*Wilsonia pusilla*). *Molecular Ecology* 11:1605-1616.

Kirchman, J.J., L.A. Whittingham y F.H. Sheldon. 2000. Relationships among Cave Swallow populations (*Petrochelidon fulva*) determined by comparisons of microsatellite and cytochrome *b* data. *Molecular Phylogenetics and Evolution* 14(1):107-121.

Knowles L.L y W.P. Maddison. 2002. statistical phylogeography. *Molecular Ecology* 11: 2623-2635.

Kumar, S., K.Tamura, IG. Jakobsen y M. Nei. 2001. MEGA2: Molecular Evolutionary Genetics Analysis software 2.1, Arizona State University, Tempe, Arizona.

Milot, E., H.L. Gibbs y K.A. Hobson. 2000. Phylogeography and genetic structures of northern populations of the yellow warbler (*Dendroica petechia*). *Molecular Ecology* 9: 667-681.

Mock, K.E., T.C.Theimer, O.E. Rhodes, D.L. Greenberg y P. Keim. 2002. Genetic variation across the historical range of the wild turkey (*Meleagris gallopavo*). *Molecular Ecology* 11:643-657.

Montes de Oca, R. 1875. Ensayo ornitológico de los troquílidos o colibríes de México. Biblioteca Nacional, Fondo Reservado. UNAM. México.

Moore, R.T. 1939. A new race of *Cynanthus latirostris* from Guanajuato. *Proceedings of the Biological Society of Washington* 53:23-29.

Moore, S.M. 1995. Inferring phylogenies from $_{MT}$DNA variation: mitochondrial-gene trees versus nuclear-gene trees. *Evolution* 49(4):718-726.

Moore, S.M. 1997. Mitochondrial-Gene trees versus Nuclear-Gene trees, a reply to Hoelzer. *Evolution*.5(12):627-629.

Moore, W.S., J.H. Graham y J.T. Price. 1991. Mitochondrial DNA variation in the northern flicker (*Colaptes auratus*, Aves). *Molecular Biology and Evolution* 8(3):327-344.

Nei, M. y S. Kumar. 2000. Molecular Evolution and Phylogenetics. Oxford University Press, Nueva York, USA

Peters, J.L. 1948. Check-list of birds of the world. Vol. VI. Harvard University Press, Cambridge.

Posada, D., K.A. Crandall y A.R. Templeton. 2000. GeoDis: a program for the cladistic nested analysis of the geographical distribution of genetic haplotypes. *Molecular Ecology* 9(4):487-488.

Pruett, C. L., S.E.Henke, S.M. Tanksley, M.F. Small, K.M. Hogan y J. Roberson. 2000. Mitochondrial DNA and morphological variation of white-winged doves in Texas. *The Condor* 102:871-880.

Ridgway, R. 1911. The birds of North and Middle America. *U.S. Natural Museum Bulletin* 50:367-377.

Rogers, A.R. 2002. The theoretical mismatch distribution. Lecture Notes on Gene Genealogies: 7: 30-46.

Rogers, A.R. y H. Harpending. 1992. Population growth makes waves in the distribution of pairwise genetic differences. *Molecular Biology and Evolution* 9:552-569.

Rozas, J. y R.Rozas. 1997. DnaSP version 2.0: a novel software package software for extensive molecular population genetic análisis. *Computer Applications in the Biosciences* 13: 307-311.

Saiton, N. y M. Nei. 1987. The neighbor-joining method: a new method for reconstructing phylogenetics. *Molecular Biology & Evolution* 4: 406-425.

Sánchez-González, L.A. 2002. Variación geográfica en *Chlorospingus ophtalmicus*. Tesis de Licenciatura. Facultad de Ciencias, UNAM.

Schneider, S., D. Roessli y L. Excoffier. 2000. Arlequin ver. 2000: a software for population genetics data analysis. Genetics and Biometry Laboratory, University of Geneva, Switzerland. URL: http://lgb.unige.ch/arlequin.

Sibley, C.G. y J.E. Ahlquist. 1990. Phylogeny and classification of birds: a study in molecular evolution.Yale University Press.

Slatkin, M. 1991. Inbreeding coefficients and coalescence times. *Genetical Research* 58:167-175.

Slatkin, M. y R.R. Hudson. 1991. Pairwise comparisons of mitochondrial DNA sequences in stable and exponentially growing populations. *Genetics* 129:555-562.

Sneath, P. y R. Sokal. 1973. Numerical Taxonomy: the principles and practice of numerical clasification. Freeman, San Franciso, California.

Swofford, D.L. 1999. PAUP* Phylogenetic analysis using parsimony, Versión 4.0. Sinnauer Associates, Sunderland MA.

Tamura, K. y M. Nei. 1993. Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpance. *Molecular Biology and Evolution* 10: 512-526.

Templeton, A. R. 1998. Nested clade analysis of phylogeographic data: testing hypotheses about gene flow and populations history. *Molecular Ecology* 7:381-397.

Templeton, A.R. 2004. Statistical phylogeography: methods of evaluating and minimizing inferences errors. *Molecular Ecology* 13:789-809.

Templeton, A.R., E. Boerwinkle y C.F. Sing. 1987. A cladisitic analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping and DNA sequence data. I. Basic theory and an analysis of alcohol dehydrogenase activity in *Drosophila*. *Genetics* 117:343-351.

Templeton, A.R., K.A. Crandall y C.F. Sing. 1992. A cladistic analysis of phenotipic associations with haplotypes inferred from restriction endonuclease mapping and DNA sequence data. III. Cladogram estimation. *Genetics* 132:619-633.

Templeton, A. R., E. Routman y C. A. Phillips. 1995. Separating population structure from population history: a cladistic analysis of the geographical distribution of mitochondrial DNA haplotypes in the Tiger Salamander *Ambystoma tigrinum*. *Genetics* 140:769-782.

Thompson, J.D., T.J. Gibson, F. Plewniak, F. Jeanmougin y D.G. Higgins. 1997. The ClustalX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Research* 24:4876-4882.

Wenink, P.W., A.J.Baker, H-U.Rösner y M.G.J. Tilanus. 1996. Global mitochondrial DNA pylogeography of holarctic breeding dunlins (*Calidris alpina*). *Evolution* 50:318-330.

Winker, K. 2000. A new subspecies of Toucanet (*Aulacorhynchus prasinus*) from Veracruz, Mexico. *Ornitología Neotropical* 11:253-257.

Wright, S. 1931. Evolution in mendelian populations. *Genetics* 16:97-159.

Wright, S. 1951 The genetical structure of populations. Ann. Eugen. 15:323-354

Zink, R.M. 1996. Comparative phylogeography in North American birds. *Evolution* 50(1): 308-317.

Zink, R.M. 1997. Phylogeographic studies of North American birds. En: Mindell, D.P. (ed.). Avian molecular evolution and systematics. Academic Press, E.U.

Zink, R.M. y R.C. Blackwell-Rago. 1997. Species limits in the Le Conte's Thrasher. *The Condor* 99:132-138.

Zink, R.M. y R.C. Blackwell-Rago. 2000. Species limits and recent population history in the Curve-billed Trasher. *The Condor* 102:881-886.

Zink, R.M. y D.L. Dittmann. 1993. Gene flow, refugia and evolution of geographic variation in the song sparrow (*Melospiza melodia*). *Evolution* 47(3): 717-729.

Zink, R.M., A.E. Kessen, T.V. Line y R.C. Blackwell-Rago. 2001. Comparative phylogeography of some aridland bird species. *The Condor* 103:1-10.

Zink, R.M. y J.V. Remsen Jr. 1986. Evolutionary processes and patterns of geographic variation in birds. *Current Ornithology* 4:1-69.

Zink, R.M., S. Rootes y D.L. Dittmann. 1991. Mitochondrial DNA variation, population structure and evolution of the common grackle (*Quiscalus quiscula*). *The Condor* 93:318-329.

Quinta parte

Las herramientas moleculares

Capítulo 16

Extracción de ácidos nucleicos

Luisa I. Falcón y Aldo Valera

Los ácidos desoxiribonucleico (ADN) y ribonucleico (ARN) son polímeros de nucleótidos formados por un azúcar de cinco carbones (desoxiribosa en el ADN y ribosa en el ARN), una base nitrogenada (que puede ser adenina, timina, citosina, guanina o uracilo) y una molécula de fosfato. El ADN es la macromolécula que contiene la información genética de las células procariontes, eucariontes y de los adenovirus. El ARN está involucrado en la síntesis de proteínas y constituye el material genético de los retrovirus (Madigan *et al.*, 2000).

La expresión de los genes es la base del metabolismo celular, y por ésta se entiende que el ADN se transcribe en ARN, que a su vez se traduce en proteínas. Esta secuencia ADN-ARN-proteínas constituye el dogma central de la biología molecular. Hay tres tipos de ARN:

- El ARN mensajero (mARN), que se sintetiza durante la transcripción gracias a la ARN polimerasa, está conformado por codones, que son secuencias de tres pares de bases que codifican un aminoácido en específico.
- El ARN de transferencia (tARN) permite el vínculo entre el tARN y los aminoácidos, ya que en un extremo cuenta con un aminoácido y en el otro, con una secuencia de tres pares de bases llamada anticodón, a la que se une el codón del mARN.
- El ARN ribosomal (rARN), que constituye la mayor parte del ARN en las células, forma junto con proteínas a los ribosomas, que facilitan el aco-

plamiento específico entre el mARN y los tARN para la traducción; los ribosomas están compuestos por tres subunidades de rARN de diferentes tamaños en los eucariontes (23S, 18S y 5S) y procariontes (23S, 16S y 5S).

Los estudios contemporáneos de ecología molecular se han esforzado en entender las funciones metabólicas de los diferentes componentes biológicos presentes en el ambiente y en conocer la diversidad biológica mediante el uso y aplicación de técnicas moleculares basadas en la amplificación de diversas regiones del ADN y ARN. La detección de la expresión de los genes o la presencia de mARN específicos nos permitirá acercarnos a temas como el impacto de los diferentes componentes físicos y biológicos del ambiente en la expresión del ADN (ya que la simple presencia de ADN no nos indica la actividad de los genes), así como a la regulación metabólica de las enzimas que son esenciales en el papel ecológico que juegan los organismos dentro de su ecosistema (ciclos biogeoquímicos, bioremediación, reciclaje de nutrientes), entre otros.

El ADN en el laboratorio

Precauciones

El trabajo de biología molecular requiere de una serie de precauciones para no contaminar las muestras con ácidos nucleicos extraños. Es necesario limpiar las áreas de trabajo con solventes como el ácido clorhídrico (10%), sosa (NaOH, 0.5 N) y/o etanol (70%). También hay que utilizar guantes para protección personal y para evitar contaminar las muestras, así como trabajar con materiales y soluciones estériles. Se recomienda trabajar en una campana de flujo laminar y si es posible, someter el material a radiación ultravioleta de 5 a 10 minutos para degradar cualquier ácido nucleico presente.

ADN

Colecta de materiales en campo

El ADN es muy estable y sólo se requiere mantener las muestras congeladas antes de su extracción. Muchos investigadores las guardan y congelan en el buffer de extracción (siguiente sección).

Métodos de extracción

En la extracción de los ácidos nucleicos, éstos se separan de cualquier otro compuesto proveniente de las células o del ambiente del cual se tomaron las muestras (p. ej. en muestras de suelos debemos considerar la alta concentración de ácidos húmicos, que son materia orgánica parcialmente degradada que no es soluble en agua a pH bajo y tiene un poder quelante muy alto, por lo que interfieren con la amplificación del ADN). Inicialmente es necesario llevar a cabo una lisis para liberar al ADN del interior celular, para lo cual se utilizan los buffers de extracción que contienen 1) detergentes, como sodio dodecilo sulfato (SDS) a una concentración final del 1% o Triton X al 0.5%; 2) una molécula quelante (p. ej. EDTA, ácido etileno amino tetra acético, 50mM), que tiene cuatro grupos carboxilo y dos grupos amino, cuya forma completamente de-protonizada puede unirse a cualquier complejo metálico en solución, quitando a los cationes de la solución para desestabilizar la membrana celular e inhibir a las ADNasas; 3) sales (p.ej. cloruro de sodio, NaCl 20mM), que forman una capa iónica suave que recubre al ADN protegiéndolo y ayudando a evitar su degradación; 4) Tris-HCl 20mM con pH entre 7.5 y 8.2 para mantener el pH de la solución estable; 5) proteinasa K (concentración final 0.5 ml ml^{-1}) para degradar proteínas o enzimas.

Durante este paso con buffer de extracción es común someter la muestra a tres o cuatro ciclos de incubación a temperaturas entre los 50°C y 70°C alternados con agitación de la muestra (manual, con vórtex o en homogenizadores celulares utilizando micro esferas) e incubación a 4 °C. La temperatura de incubación no puede ser mayor a los 80°C, ya que a esta temperatura se comienza a degradar el ADN. Estas incubaciones facilitan la ruptura de lípidos en la membrana celular, permitiendo la liberación del ADN de la estructura celular.

Posteriormente, resulta necesario tratar las muestras con dos extracciones de fenol para eliminar las proteínas (como nucleasas); para ello se agrega de fenol el mismo volumen de la muestra, se mezcla bien (vórtex), centrifuga a 10,000 xg (tres minutos), recuperando el sobrenadante con cuidado de no acarrear las proteínas que quedan en la interfase. Hay que considerar que no todos los paquetes comerciales de extracción llevan a cabo este paso, por lo que se podría observar una degradación de las muestras por nucleasas con el tiempo, aun cuando se guarden en el congelador. Después de la extracción con fenol, es común llevar a cabo otra extracción con cloroformo para acabar de limpiar la muestra de cualquier residuo lipídico. Una vez más, hay que agregar igual volumen de cloroformo que el de la muestra, mezclar bien (vórtex)

y centrifugar para recuperar el sobrenadante. Una vez concluido este paso, se lleva a cabo la precipitación del ADN con etanol absoluto y con una sal a alta concentración (p.ej. acetato de sodio a una concentración final de 0.3 M; cloruro de sodio a una concentración final de 0.2 M; acetato de amonio a una concentración final de 2.5 M o cloruro de litio, concentración final 0.8 M). Por otra parte, con la adición de la sal, el ADN que está cargado negativamente va a obtener una capa iónica positiva que permite su precipitación. La elección de qué sal usar depende de lo que se quiera hacer con el ADN después de su extracción; por ejemplo, no se debe utilizar precipitación con acetato de sodio si el ADN va a ser fosforilado con una quinasa polinucleotídica; el cloruro de sodio puede ser usado si hay SDS en la solución ya que esta sal permite que el SDS permanezca en solución en presencia de etanol.

Por otra parte, el alcohol en la solución de precipitación se utiliza para remover la concentración residual de sales y promover la precipitación del ácido nucleico. Cuando se trata de muestras con baja concentración de ácidos nucleicos, comúnmente se utiliza un volumen de muestra de isopropanol, en lugar de dos volúmenes de muestra cuando se trabaja con etanol. La precipitación del ADN es casi inmediata en presencia de la sal y el alcohol, sin embargo se recomienda incubar la muestra durante 20 minutos a -80ºC o durante 45 minutos a -20ºC. Posteriormente, se centrifuga la muestra (15 minutos a 10,000 xg), se remueve la fase acuosa y se lava la pastilla de ADN con etanol al 70% para eliminar todas las sales que permanezcan en la solución. Las muestras se vuelven a centrifugar (un minuto a 10,000 x g), se elimina el etanol y se deja secar el ADN (15 minutos a 37º C o media hora a temperatura ambiente) hasta que no haya mas trazas de alcohol.

Protocolos de extracción de ADN implementados en el laboratorio de Evolución Molecular y Experimental del Instituto de Ecología

En el laboratorio se han utilizado tres tipos de métodos para extracción de ADN, que se han modificado de protocolos previos propuestos por algunos autores como Doyle y Doyle (1987):

- Extracción de ADN en bacterias gram negativas
- Extracción de ADN en plantas
- Extracción de ADN en animales

Estos métodos se caracterizan por no requerir en ningún momento del uso de paquetes comerciales, las soluciones requeridas son elaboradas en el laboratorio y pueden ser modificadas según requerimientos particulares. La extracción de ADN por estos métodos combina procesos químicos, físicos y mecánicos e incluye básicamente tres pasos:

1. Lisis celular
2. Eliminación de proteínas
3. Precipitación y limpieza del ADN

Nuestros métodos de extracción de ADN presentan algunas ventajas y desventajas respecto a otros métodos: son económicos, no se manejan compuestos químicos tóxicos o contaminantes al ambiente y se obtienen grandes cantidades de ADN; pero por otro lado, la pureza del ADN no es absoluta y en algunas especies se obtiene un ADN parcialmente degradado (en pedazos).

Paralelamente, también se han usado paquetes comerciales cuando se requiere ADN de alta calidad o bien se trata de organismos de cierto grado de complejidad como algunas bacterias gram positivas o plantas que naturalmente producen gran cantidad de compuestos secundarios. Algunos paquetes comerciales de alto rendimiento son:

- PURIGEN
- QUIAGEN
- MILIPORE
- GENTRA
- MoBio

En algunos casos hemos modificado los protocolos de uso de dichos paquetes, esto con el fin de ajustarlos a los requerimientos de los organismos en estudio o para mejorar los resultados obtenidos.

Extracción de ADN de bacterias gram negativas

1. Obtener un cultivo nuevo de bacterias. En un microtubo de 1.5 ml agregar 0.5 ml de sulfato de magnesio 10 mM y resuspender en él una asada de bacterias.

2. Centrifugar a 4500 xg por 3 minutos y eliminar el sobrenadante perfectamente escurriendo las muestras en papel secante.
3. Resuspender en 300 µl de buffer Tris HCl_2 0.1M pH 8.
4. Adicionar 40 µl de lisozima (20 mg/ml) y 5µl de ARNasa (7000 U/ml).
5. Incubar a 37° C por 5 min., posteriormente a -20°C durante 10 min. y finalmente a 85° C durante 10 min. Entre cada incubación, mezclar vigorosamente.
6. Agregar 20 µl de proteinasa K (20 mg/ml).
7. Incubar a 70°C durante 10 min., durante la incubación invertir manualmente 5-10 veces.
8. Centrifugar a 8 000 xg durante 1 min.
9. Recuperar el sobrenadante (aprox. 250 µl), sin tocar el botón y transferir a un tubo nuevo. Adicionar 150 µl de isopropanol absoluto frío (a -20° C).
10. Incubar a -20° C durante 30 min.
11. Centrifugar a 10 000 xg durante 3 min. Eliminar perfectamente el sobrenadante.
12. Limpiar el botón con 300 µl de etanol al 70% agitando brevemente en vórtex.
13. Centrifugar a 10 000 xg por 3 min. Eliminar perfectamente el sobrenadante.
14. Secar perfectamente el botón a 65° C por 5 min. y resuspender en 50 a 00µl de agua pura, desionizada y estéril.

Un método alternativo y muy eficiente es utilizar las soluciones del paquete de la marca PUREGENE, al que aplicamos las siguientes modificaciones:

Tratamiento previo

1. Cultivar las bacterias en medio sólido, pobre en nutrientes, durante el tiempo necesario para obtener un crecimiento abundante pero de células jóvenes.
2. En un tubo de 1.5 ml estéril agregar 0.5 ml de sulfato de magnesio 10 mM.
3. Agregar una asada pequeña de bacterias (calculando una cantidad de 2 colonias medianas) y homogeneizar perfectamente "pipeteando" suavemente.
4. Incubar en hielo durante 10 min.
5. Centrifugar a 7000 xg durante 5 min. a 4° C.
6. Eliminar totalmente el sobrenadante colocando los tubos boca-abajo en un papel secante.

Lisis celular

1. Agregar 300 µl de solución de lisis y resuspender las células pipeteando suavemente varias veces.
2. Incubar a 80°C por 5 min. Es conveniente mezclar invirtiendo los tubos 5-10 veces.
3. Incubar a –20°C por 5 min.
4. Mezclar vigorosamente durante 10 seg.

Tratamiento con ARNasa

1. Agregar 4 µl de solución de ARNasa.
2. Mezclar invirtiendo los tubos 25-30 veces.
3. Incubar a 37°C durante 15-30 min. Conservar en hielo.

Precipitación de proteínas

1. Agregar 100 µl de Solución de Precipitación.
2. Mezclar vigorosamente durante 20 seg.
3. Centrifugar a 10 000 xg por 3 min.
4. Tomar 300 µl del sobrenadante, cuidando de no tocar el botón y transferirlos a un tubo nuevo. Conservar en hielo.

Precipitación de ADN

1. Agregar 300 µl de isopropanol absoluto y frío (a -20°C).
2. Mezclar invirtiendo los tubos 50 veces. Dejar reposar durante 1 hora a -20°C.
3. Centrifugar a 8 000 xg durante 1-2 min.
4. Eliminar el sobrenadante y escurrir perfectamente los tubos en papel secante.
5. Agregar 300 µl de etanol frío al 70%. Mezclar vigorosamente durante 5 seg.
6. Centrifugar a 10 000 xg y eliminar perfectamente el sobrenadante escurriendo en papel secante.
7. Incubar el botón a 65°C durante 1 min. ventilar en la campana de flujo laminar por 5 min.

Hidratación del ADN

1. Agregar de 20 a 50 µl de Solución de Hidratación.
2. Incubar a 65°C durante 5min.

Extracción de ADN en plantas

Para la extracción de ADN en plantas hemos usado con éxito los siguientes protocolos, que son modificados del protocolo original conocido como *mini-prep* (Doyle y Doyle, 1987). Con estos protocolos hemos extraído ADN de varios grupos de plantas como agaváceas, coníferas, crasuláceas, cactáceas, leguminosas, entre otras. El protocolo básico es el siguiente:

Mini-prep

1. En un mortero moler alrededor de 1 g de tejido con nitrógeno líquido hasta obtener un polvo fino.
2. Agregar 1 ml de buffer CTAB 2X y seguir moliendo. Recuperar en un microtubo de 1.5 ml.
3. Centrifugar a 8 000 xg durante 8 min. a 4°C.
4. Eliminar el sobrenadante y resuspender con 600 µl de CTAB 2X.
5. Incubar a 60ºC durante 10 min.
6. Agregar 600 µl de cloroformo:octanol 24:1, agitar hasta homogeneizar.
7. Centrifugar a 5 000 xg durante 12 min. a 4°C (o hasta que le sobrenadante quede transparente).
8. Trasladar el sobrenadante a un tubo nuevo, cuidando de no tomar la interfase.
9. Agregar 2/3 del volumen final de isopropanol frío para precipitar el ADN.
10. Dejar reposar durante la noche a -20°C.
11. Centrifugar a 8000 xg durante 5 min a 4°C. Eliminar perfectamente el sobrenadante.
12. Limpiar el ADN agregando 1 ml de etanol 70% frío y centrifugar a 7 000 xg durante 5 min.
13. Eliminar el sobrenadante y resuspender con 100 µl de buffer TE.

Este protocolo es eficiente para la extracción de ADN de plantas que **no** producen compuestos secundarios o gran cantidad de polisacáridos; en casos

contrarios el protocolo se ha modificado en función de los requerimientos de cada especie, básicamente en los tiempos y cantidades de las soluciones utilizadas. Así, tenemos que para la mayoría de las especies del género *Agave* el protocolo ha sido modificado como sigue:

1. En un mortero cortar en cuadros pequeños alrededor de 1 g de tejido y moler con nitrógeno líquido hasta obtener un polvo fino, eliminar el máximo de fibras. De ser posible, tratar de moler en un solo evento y no remoler la muestra excesivamente (esto provoca que el ADN se fraccione).
2. Agregar 250 µl de buffer de extracción CTAB y 750 µl de buffer STE, seguir moliendo y recuperar en un microtubo de 1.5 ml (es importante conservar las proporciones de buffer, sin embargo las cantidades pueden variar según la textura y consistencia de la muestra tratando de obtener un jarabe).
3. Centrifugar a 9000 xg durante 8 min. y eliminar el sobrenadante. Si el sobrenadante resulta de un color intenso, se recomienda repetir los pasos 2 y 3.
4. Resuspender con 400 µl de CTAB 2X y 600 µl de STE.
5. Centrifugar a 9 000 xg durante 8 min.
6. Eliminar el sobrenadante y resuspender con 600 µl de CTAB 2X.
7. Agregar 4 µl de ARNasa (7000 U/ml). Incubar 20 min a 37º C. Conservar en hielo 5 minutos.
8. Agregar 40 µl de proteinasa K e incubar a 65ºC por 20 minutos. Durante la incubación es conveniente invertir los tubos 2 a 3 veces. Poner en hielo 5 minutos.
9. Agregar a cada tubo 600 µl de cloroformo:octanol 24:1, agitar en vortex hasta homogeneizar (no exceder la agitación).
10. Centrifugar a 8 000 xg durante 12 min. (o hasta que el sobrenadante quede transparente, o bien repetir el paso de cloroformo:octanol con el sobrenadante resultante del primer centrifugado).
11. Trasladar el sobrenadante a un tubo nuevo. Este paso es fundamental para obtener un ADN limpio, se debe evitar recuperar material de la difase que se forma; recupere con una pipeta de 200 µl sólo el 70% del sobrenadante: tendrá menos ADN pero más limpio.
12. Precipitar el ADN agregando isopropanol frío a -20°C en un volumen de 2/3 de lo obtenido de sobrenadante en el paso anterior (esto es, 600 µl aprox.).

13. Dejar reposar 2 horas a -20ºC. No es recomendable dejar precipitando muchas horas pues el producto resulta más sucio.
14. Centrifugar a 7 500 xg durante 5 min. Eliminar el sobrenadante. Se observará un pequeño botón al fondo del tubo, que debe ser color blanco traslúcido. Si es muy blanco y muy grande es posible que esté sucio; es conveniente eliminar muestras de este tipo.
15. Limpiar el ADN agregando 1 ml de etanol 70% frío a -20°C, mezclar en vortex hasta observar que el botón se desprende. Dejar reposar 5 minutos a temperatura ambiente y centrifugar a 9 000 xg durante 5 min.
16. Eliminar el sobrenadante sin perder de vista el botón. Si lo considera necesario se puede repetir el lavado con alcohol al 80%.
17. Rehidratar el ADN con agua ultra pura (en nuestro laboratorio usamos agua Mobio con no. de cat.: 17012-5200), el volumen de agua depende del tamaño del botón y puede ir de 15 µl a 150 µl.

Reactivos:

Buffer de extracción CTAB 2X

Tris-HCl 100 mM pH8
NaCl 1.4 M
EDTA 20 mM pH8
CTAB 2%
b-mercaptoetanol 0.3% (no agregar este reactivo hasta el momento de usar el buffer)

Buffer de extracción CTAB

Tris-HCl 100 mM pH8
NaCl 1.5 M
EDTA 20 mM pH8
CTAB 4%
PVP40 4%
Ac. ascórbico 0.1%
DIECA 0.1%
b-mercaptoetanol 0.3% (no agregar este reactivo hasta el momento de usar el buffer)

Buffer de extracción STE

Tris-HCl 100 mM pH8
EDTA 50 mM pH8
NaCl 100mM
b-mercaptoetanol 0.3% (no agregar este reactivo hasta el momento de usar el buffer)

Nota importante: Un factor determinante para obtener ADN de calidad es controlar la concentración de sales residuales, que deben ser eliminadas. Para ello existen protocolos de limpieza con fenol-cloroformo que se presentan más adelante, pero nosotros hemos llegado a controlar el exceso de sales directamente desde la elaboración de los buffers de extracción, evitando con ello el manejo del fenol que es altamente contaminante y peligroso. Lo que hacemos realizar una extracción preliminar y medir su calidad en un *biofotómetro*; si las lecturas indican exceso de sales reducimos el NaCl al mínimo posible. Esto se debe a que algunas plantas acumulan sales en sus tejidos y la reacción se satura de sales.

En otros grupos de plantas como cactáceas y crasuláceas, la extracción de ADN resulta más difícil por la presencia de una gran cantidad de mucopolisacáridos y compuestos secundarios, en cuyo caso hemos logrado la extracción aplicando el siguiente protocolo básico, que modificamos en algunos pasos según los requerimientos de las plantas.

Extracción de ADN total de plantas suculentas por el método CTAB-STE

1. Moler 0.5 g de tejido fresco con nitrógeno líquido hasta obtener un polvo fino. De ser posible, tratar de moler en un solo evento y no remoler la muestra excesivamente (esto provoca que el ADN se fraccione).
2. Agregar 260 µl de CTAB y 975 µl de STE, agitando hasta obtener un jarabe.
3. Agregar 65 µl de SDS al 20% agitando vigorosamente por 5 minutos e incubar a 65°C por 10 minutos.
4. Agregar 325 µl de acetato de potasio 5M frío e incubar a -20°C por 40 minutos.
5. Centrifugar los tubos a 12 000 xg por 30 minutos.
6. Filtrar la fase acuosa con un trozo de algodón estéril insertado en la boca del tubo (no muy apretado ni muy suelto) y sobre éste recuperar el sobrenadante en un tubo nuevo.

7. Agregar 2/3 del volumen de sobrenadante obtenido con isopropanol frío (-20°C), agitar suavemente e incubar a -20°C por 1 hr.
8. Mezclar suevemente, y centrifugar los tubos a 12 000 xg a 6°C por 10 minutos. Eliminar el sobrenadante, dejar secar el botón de ADN y resuspender en 0.5 ml de agua ultra pura. Incubar a 65°C por 10 min.
9. Centrifugar a 12 000 xg por 10 minutos y transferir el sobrenadante a un tubo nuevo.
10. Agregar 30 µl de acetato de sodio 3M y 300 µl de isopropanol frío (-20°C) y mezclar suavemente. Incubar a -20°C por 30 minutos.
11. Mezclar suavemente y centrifugar a 6°C por 10 minutos, se formará un botón pequeño traslúcido.
12. Limpiar el botón con etanol frío (-20°C) al 80 %, mezclando en vortex hasta soltar el botón. Si lo considera necesario se puede repetir el lavado con alcohol al 80%.
13. Centrifugar a 1000 x g y eliminar el sobrenadante.
14. Rehidratar el ADN con agua ultra pura (en nuestro laboratorio usamos agua Mobio con no. de cat.: 17012-5200), el volumen de agua depende del tamaño del botón y puede ir de 15 µl a 150 µl.

Reactivos:

Buffer de extracción CTAB

Tris-HCl 100 mM pH8
NaCl 1.5 M
EDTA 20 mM pH8
CTAB 4%
PVP40 4%
Ac. Ascorbico 0.1%
DIECA 0.1%
b-mercaptoetanol 0.3% (no agregar este reactivo hasta el momento de usar el buffer)

Buffer de extracción STE

Tris-HCl 100 mM pH8
EDTA 50 mM pH8
NaCl 100mM

b-mercaptoetanol 0.3% (no agregar este reactivo hasta el momento de usar el buffer)

SDS 20%
Acetato de potasio 5 M
Acetato de sodio 3 M

En algunas ocasiones nos hemos enfrentado a muestras verdaderamente complicadas como *Opuntia rastrera*, en cuyos casos iniciamos el proceso con una maceración en alcohol etílico al 96%, lo que rompe los mucopolisacáridos y nos permite seguir con los procedimientos normales. En ocasiones también combinamos partes de un protocolo con otro para resolver problemas específicos. Cuando es necesario limpiar el ADN porque se encuentra muy saturado de sales o proteínas que impiden la amplificación de productos de PCR utilizamos el siguiente método:

Limpieza de las muestras con fenol-cloroformo

1. Agregar 50 µl de fenol y 50 µl de cloroformo y agitar en el vórtex para mezclar perfectamente.
2. Centrifugar a 10 000 xg por 30 segundos.
3. Rescatar la fase superior, con cuidado, sin tocar la interfase blanca que contiene las proteínas y ponerla en otro tubo limpio.
4. Agregar 50 µl de fenol y 100 µl de cloroformo, agitar en vórtex y centrifugar en microcentrífuga como se hizo en los pasos anteriores.
5. Repetir estos pasos hasta que no se vea turbia la fase superior, y cuando esto suceda ponerle 100 µl de coloroformo, mezclar bien en el vórtex y centrifugar. Rescatar la fase superior a un tubo nuevo.
6. Precipitar el ADN agregando 10 µl de NaCl 2M y 250 µl de isopropanol y reposar la muestra a -20°C 4 hr. máximo.
7. Centrifugar los tubos por 1 hora a 12 000 x g a 6°C, eliminar el sobrenadante y lavar el precipitado con 1 ml de etanol al 80% frío.
8. Secar el precipitado y resuspender en 20-40 µl agua ultra pura.

Extracción de ADN en animales

Para la extracción de este material hemos utilizado el protocolo mini-prep modificado básicamente con un aumento de proteinasa K y de ARNasa. Hemos

mantenido semejante el resto del protocolo aun cuando en el caso de insectos hemos requerido cantidades mínimas de tejido, como 0.1 g. También hemos combinado partes del protocolo mini-prep con el uso de paquetes comerciales para células animales, como las células hepáticas de borrego cimarrón, *Ovis canadensis*, de las que se obtuvo ADN de buena calidad.

Cuando no obtenemos la calidad requerida para los experimentos a realizar, recurrimos al uso de los paquetes comerciales mencionados, ya que a pesar de ser muy caros y no permitir realizar estudios extensivos, permiten obtener ADN de buena calidad. Respecto a los costos de dichos paquetes, basta decir que siempre es más barato no tener que repetir todo el experimento y que es posible obtener buenas promociones si se consiguen los componentes a granel y en grandes volúmenes, y no en paquete, pues siempre es necesario ajustar los protocolos.

ARN

Precauciones

Una vez colectado el material es necesario agregarle una solución que estabilice al ARN lo antes posible, ya que los cambios en el patrón de expresión de los genes debido a la degradación del ARN o a inducciones en la transcripción ocurren inmediatamente. Debido a su estructura química el ARN es una molécula muy frágil que puede romperse por acción de los grupos 2'-OH (altamente reactivos) adyacentes al esqueleto de ribosa-fosfato. Existen paquetes comerciales para estabilización de muestras de ARN de compañías como QIAGEN, Ambion y RNA-works, que en general funcionan con una mezcla de anticuerpos específicos para ribonucleasas, inhibidores de ribonucleasas y/o contienen compuestos que neutralizan a los grupos 2'-OH (como es el caso de RNA-works).

Dado que las ARNasas no van a ser inactivadas en su totalidad en el autoclave, resulta esencial trabajar con material estéril y de preferencia de un solo uso. El material de vidrio debe lavarse con una solución 0.5 N de NaOH, enjuagarse tres veces con agua de-ionizada y meterse al autoclave durante 2 horas. El material de plástico deberá limpiarse con etanol absoluto y meterse al autoclave durante 45 minutos. Las micro esferas de vidrio, generalmente utilizadas en la extracción para incrementar la lisis celular por acción física, deben someterse a combustión durante 12 hrs. a 450º C y deben pasar en el autoclave 45 minutos.

Es importante que los reactivos que se utilicen en la extracción de ARN se mantengan en alícuotas con los volúmenes mínimos necesarios para cada etapa del proceso y en caso de ser posible se pasen dos veces por el autoclave.

Cuando se extraen los ácidos nucleicos para su amplificación, es indispensable separar los sitios en los que se lleva a cabo la extracción y la amplificación, o en su caso, limpiar a profundidad la zona donde se llevó a cabo la extracción con NaOH 0.5N y con etanol absoluto antes de amplificar. En cada una de estas dos etapas es ideal utilizar juegos de pipetas diferentes, o en su defecto usar puntas de pipeta con filtro en la preparación de las mezclas de reactivos a utilizar en la amplificación. También resulta propicio trabajar siempre bajo la campana de flujo laminar con dos mecheros encendidos.

Métodos de extracción del ARN

GIPS

El método utilizado con más frecuencia es el de GIPS (por sus siglas en inglés; acido tiosanato de guanidina, fenol, sarcosyl) desarrollado por Chomczynski y Sacchi (1987). Cada uno de estos componente juega un papel específico: el ácido tiosanato de guanidina es sumamente fuerte y tiene un alto poder denaturalizante; el fenol, al estar acidificado, provoca que el ADN se acumule en la interfase entre la fase acuosa y la del fenol, dejando al ARN en la fase acuosa; el sarcosyl, por su parte, es un detergente muy potente que ayuda a la lisis celular, pero no reemplaza la ruptura física de las células que generalmente se logra con micro esferas de vidrio agitadas con la ayuda de un vórtex o de un homogenizador celular (como Bead-Beater). Sin embargo, es importante tratar las muestras con ADNasas libres de ARNasas.

BOOM

El método BOOM, desarrollado por Boom *et al.* (1990) lleva a cabo la extracción con el ácido tiosanato de guanidina, separando los ácidos nucleicos con base en su alta afinidad para enlazarse en matrices de sílica, en vez de utilizar fenol. Dado que este método aisla tanto ADN como ARN, resulta esencial utilizar las nucleasas específicas para ADN que lo eliminen de la muestra.

Algunas compañías han creado paquetes de extracción basados en este principio, que utilizan matrices de sílica diseñadas para favorecer el enlace de ARN (como RNeasy de QIAGEN).

Las moléculas grandes de ARN, incluyendo ARNr y ARNm, son insolubles en soluciones con altas concentraciones de sal, por lo que durante su precipitación es común utilizar cloruro de litio 8 M (libre de ARNasas) e incubar

las muestras a 0ºC durante dos horas antes de centrifugar. Sin embargo, este método no debe ser utilizado para ARN que será sometido a transcripción reversa (véase más adelante).

Transcripción reversa

Esta técnica permite pasar, en sentido inverso al dogma central de la biología molecular, de ARN a ADN. El primer paso es convertir el mARN en ADN complementario (cADN) a partir de la actividad de la reverso transcriptasa. Esta enzima existe en la naturaleza como parte del mecanismo de replicación en virus, y utiliza un tARN como oligonucleótido o *primer*.

Método utilizado para transcripción reversa para expresión de genes involucrados en la fijación de nitrógeno en muestras acuáticas (modificado por Zani *et al.*, 2000 del kit de Promega de RT-PCR)

- Agua ultra pura sin nucleasas (28 µl)
- Buffer de virus de mieloblastosis de ave 5x (10 µl)
- dNTP, a concentración de 10 mM por núcleotido) (1 µl)
- Primer a 12.5 M (2 µl)
- Reverso transcriptasa de virus de mieloblastosis de ave (1 µl)
- Muestra de ARN libre de ADN (1 µl)
- Ciclo
- 30 min a 42° C

El cADN formado durante la transcripción reversa es posteriormente utilizado en la reacción de PCR para amplificarlo e identificar el gen que estaba originalmente expresándose en nuestra muestra (ver capítulo 17).

Método para aislamiento de ARN total de muestras de agua ambientales (modificado a partir del kit de QIAGEN, Rneasy; Falcón et al., 2004)

1 Colectar muestras de agua en filtros DURAPORE (Millipore, Corp. San José, CA. EUA).
2 Colocar filtros con las células hacia fuera dentro de tubos Eppendorf de 1.5 ml con micro esferas de vidrio y con la solución RLT que contiene

ácido tiosanato de guanidina y β-mercaptoetanol (que desactiva moléculas complejas y ayuda en la lisis celular).

3 Aplicar vibración con vórtex o con homogenizador celular (vibración 1 min, colocar en hielo 1 min, vibración 1 min).

4 Centrifugar (2 min x 10000 xg) y transferir el supernadante (con ARN) a tubo colector nuevo. Agregar 350 µl de etanol absoluto para limpiar y ayudar a la precipitación del ARN. Mezclar bien por pipeteo o vórtex.

5 Transferir mezcla a tubo de centrífuga de 1.5 ml con filtro de sílica (columna); colocar columna en tubo colector. Centrifugar (15 seg x 10000 x g) para atrapar el ARN en la matriz de sílica.

6 *En caso de utilizar ADNasa, agregar 350 µl de la solución RW1, 80 µl de la solución de ADNasa (10 µl de la solución stock -1500 unidades de ADNasa sólida disuelta en 550 µl de agua ultra pura libre de nucleasas- en solución RDD). Centrifugar (15 seg x 10000 xg). Repetir la limpieza con solución RW1 (350 µl) y centrifugar (15 seg x 10000 xg). Si no se utiliza ADNasa pasar directamente al siguiente paso.

7 Limpiar la muestra con la solución RW1 (que contiene ácido tiosanato de guanidina). Agregar 700 µl y centrifugar 15 seg x 10000 xg.

8 Colocar la columna en tubo colector nuevo y limpiar con 500 µl de solución RPE (que contiene etanol absoluto); centrifugar 15 seg 10000 xg. Repetir.

9 Centrifugar 1 min x 10000 xg para limpiar la muestra de todo etanol residual.

10 Colocar columna en tubo colector nuevo y agregar entre 30 y 50 µl de agua ultra pura libre de nucleasas. Esperar un minuto para separar las hebras de ARN de la membrana de sílica y centrifugar (1 min x 10000 xg).

11 Guardar el ARN aislado (total) en nitrógeno líquido o a -80º C.

Bibliografía

Boom, R., C J Sol, M M Salimans, C L Jansen, P M Wertheim-van Dillen y J van der Noordaa, 1990. Rapid and simple method for purification of nucleic acids. *Journal of Clinical Microbiology* 28(3): 495-503.

Chomczynski, P. y N Sacchi, 1987. RNA isolation from cultured cells. *Analytical Biochemistry* 162: 156-159.

Doyle, JJ y JL Doyle, 1987.A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochemistry* Bulletin 19: 11-15.

Falcón, L. I., E. J. Carpenter, F. Cipriano, B. Bergman y D. G. Capone. 2004. N2-fixation by unicellular bacterioplankton in the Atlantic and Pacific Oceans: phylogeny and in situ rates. *Applied Environmental Microbiology* 70: 765-770.

Madigan, MT., JM Martinko y J Parker, 2000. *Brock biology of microorganisms*. Prentice Hall, Upper Saddle River, NJ. 991 pp.

Zani, S., M.T. Mellon, J.L. Collier y J.P. Zehr, 2000. Expression of nifH genes in natural microbial assemblages of Lake George, NY detected with RT-PCR. *Applied Environmental Microbiology* 66: 3119-3124.

Capítulo 17

Guía práctica sobre la técnica de PCR

Laura Espinosa Asuar

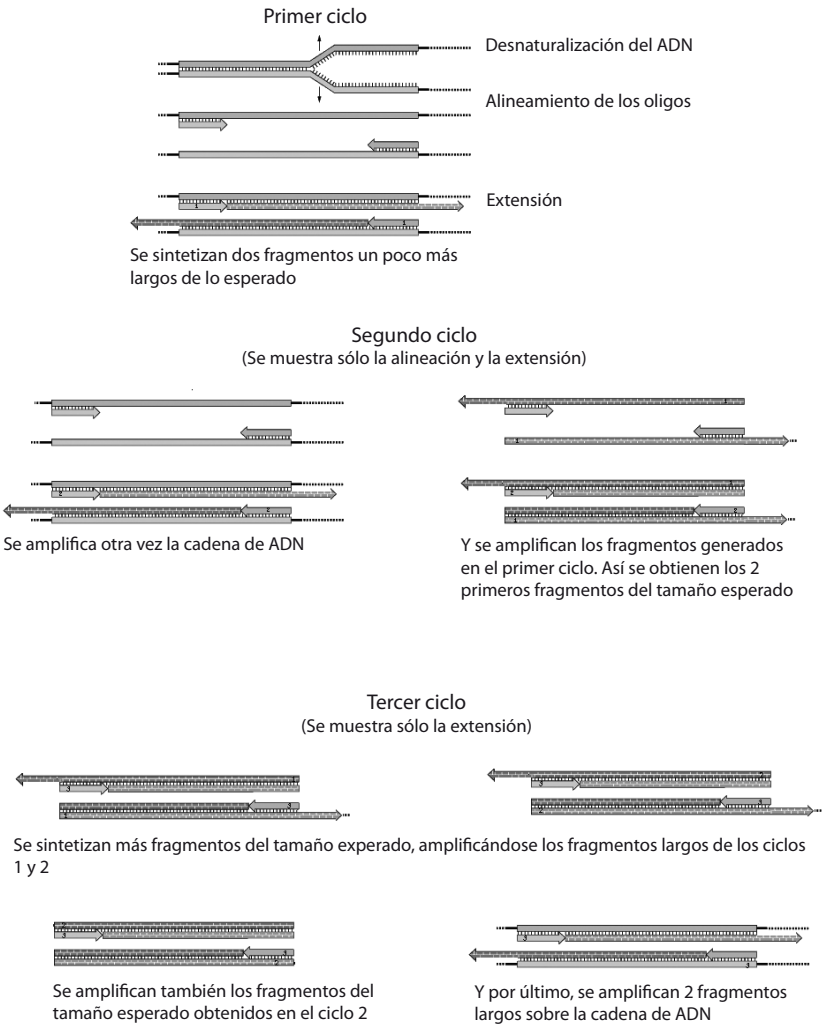
PCR son las siglas en inglés de *Polymerase Chain Reaction* o Reacción en Cadena de la Polimerasa. La idea básica de la técnica es sintetizar muchas veces un pedazo o fragmento de ADN utilizando una polimerasa que puede trabajar a temperaturas muy elevadas, ya que proviene de la bacteria *Thermus aquaticus* que vive a altas temperaturas (79ºC a 85ºC), de ahí su nombre comercial más conocido: *taq* polimerasa. Cuando hacemos una reacción de PCR simulamos lo que sucede en una célula cuando se sintetiza el ADN y en el tubo se mezclan todos los ingredientes necesarios para hacerlo: la polimerasa, el ADN del organismo que queremos estudiar –donde se encuentra el fragmento que queremos sintetizar– , los oligonucleótidos (llamados también *primers*, iniciadores, cebadores, "oligos", etc.) necesarios para que se inicie la transcripción, dinucleótidos (dNTPs), y las condiciones para que la enzima trabaje adecuadamente (cierto pH, determinadas cantidades de magnesio en forma de $MgCl_2$, KCl, y pueden necesitarse otras sales o reactivos, dependiendo de cada polimerasa). Esta técnica tan ingeniosa tiene muchísimas aplicaciones distintas y se ha convertido en una herramienta muy importante en la biología molecular; sus aplicaciones van desde la genética de poblaciones, evolución molecular y genómica hasta la medicina forense.

¿Pero cómo funciona el PCR? Supongamos que ya tenemos los tubos listos con todo lo necesario para que la síntesis del fragmento que nos interesa que se lleve a cabo (*taq* polimerasa, dinucleótidos, ADN, agua, buffer con mag-

nesio y otras sales, y oligonucleótidos). El siguiente paso es colocar los tubos en una máquina conocida como termociclador, que básicamente sirve para calentarlos o enfriarlos a temperaturas muy precisas. ¿Cómo es que se amplifica el (o los) fragmento(s) que queremos? Primero, para hacer más sencilla la explicación, vamos a suponer que esperamos un solo fragmento de un tamaño determinado, y lo que sucede es lo siguiente (ver el primer ciclo de la figura 1): el termociclador calienta o enfría los tubos a tres temperaturas distintas, que se repiten una y otra vez (lo que se llama los ciclos de reacción), la primera es a 95ºC (y a este paso se le llama desnaturalización) durante la cual las dobles cadenas del ADN se abren o desnaturalizan, quedando en forma de cadenas sencillas; después el termociclador ajusta la temperatura en un intervalo entre 40º y 60ºC (llamada de alineamiento), a esta temperatura se forman y se rompen constantemente los puentes de hidrógeno entre los oligonucleótidos y el ADN, y aquellas uniones más estables (las que son complementarias) durarán mayor tiempo, quedando los oligonucleótidos "alineados" formando una pequeña región de doble cadena. La polimerasa se une a este pequeño pedazo de ADN de doble cadena y comienza a copiar en sentido 5' a 3'; al agregar unas bases más, los puentes de hidrógeno que se forman entre las bases estabilizan más la unión y el oligonucleótido permanece en este sitio para el siguiente paso. Después la temperatura sube a 72ºC (paso que se conoce como extensión), ya que 72ºC es la temperatura en la cual la polimerasa alcanza su máxima actividad, y continúa la síntesis de los fragmentos de ADN a partir de los oligonucleótidos que ya se habían alineado.

En el primer ciclo, con estas tres temperaturas, se sintetizarán los primeros fragmentos a partir del ADN genómico. Estos primeros fragmentos no tendrán el tamaño esperado, serán un poco más grandes ya que la *taq* copiará hasta donde le sea posible, pero como veremos más adelante, se obtendrán en cantidades tan pequeñas que al final no podremos detectarlos. Después se repiten una vez más las tres temperaturas, pero en este segundo ciclo, los oligonucleótidos, además de unirse al ADN que pusimos al inicio, también se unirán a los fragmentos recién sintetizados del primer ciclo (ver segundo ciclo de la figura 1), por lo tanto en este segundo paso la polimerasa sintetizará 2 fragmentos largos copiados directamente del ADN y 2 fragmentos del tamaño esperado, que es el tamaño que hay entre los dos oligonucleótidos que hemos usado. De esta forma con cada ciclo aumentará el número de fragmentos del tamaño que queremos. Cabe mencionar que antes y después de estos ciclos se programan dos pasos, uno de 95ºC durante varios minutos para iniciar con desnaturalización, y al final de los ciclos, un paso último de extensión a

Figura 1. Descripción del proceso de un PCR en los primeros ciclos de reacción. Comúnmente se utilizan 35 ciclos para amplificar el fragmento que se requiere (modificada de Parkes, 2003)


Primer ciclo
Desnaturalización del ADN
Alineamiento de los oligos
Extensión
Se sintetizan dos fragmentos un poco más largos de lo esperado
Segundo ciclo
(Se muestra sólo la alineación y la extensión)
Se amplifica otra vez la cadena de ADN
Y se amplifican los fragmentos generados en el primer ciclo. Así se obtienen los 2 primeros fragmentos del tamaño esperado
Tercer ciclo
(Se muestra sólo la extensión)
Se sintetizan más fragmentos del tamaño experado, amplificándose los fragmentos largos de los ciclos 1 y 2
Se amplifican también los fragmentos del tamaño esperado obtenidos en el ciclo 2
Y por último, se amplifican 2 fragmentos largos sobre la cadena de ADN

Fuente: Parkes H. 2003. Food for Thought http://www.chemsoc.org/chembytes/ezine/1999/parkes_may99.htm, accesado 08/03.

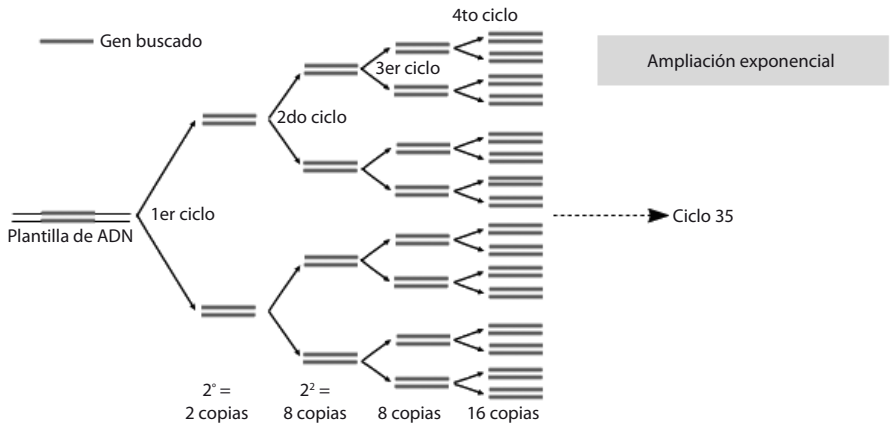
72ºC para permitir que la *taq* termine de sintetizar todos los fragmentos que pueden haber quedado incompletos.

Para este tipo de PCR es necesario que uno de los oligonucleótidos tenga la misma secuencia que se encuentra en una de las cadenas del ADN, y el otro deberá llevar la secuencia complementaria que estará al final del fragmento que se quiere amplificar (por lo cual se les llama *forward* y *reverse*) para que uno sea complementario a la cadena que forma el otro; si no es así no podría amplificarse el sitio que se necesita. Como cada pedazo sintetizado sirve como base para sintetizar otros en el siguiente ciclo, el número de copias aumentará en forma exponencial (ver tercer ciclo de las figuras 1 y 2). Con una sola molécula de ADN, en el ciclo 1 se producen 2^1=2 nuevos fragmentos, en el ciclo 2 serán 2^2, esto es, 4 fragmentos recién sintetizados, y así, con 35 ciclos de PCR se producirán $2^1+2^2\ldots+2^{34}+2^{35}=2^{36}$ nuevos fragmentos, de los cuales sólo 70 serán fragmentos de un tamaño mayor al esperado (2 por cada ciclo) obtenidos al sintetizarlos directamente del ADN genómico; esta pequeña cantidad es casi imposible de detectar al analizar nuestros productos.

Tipos de técnicas

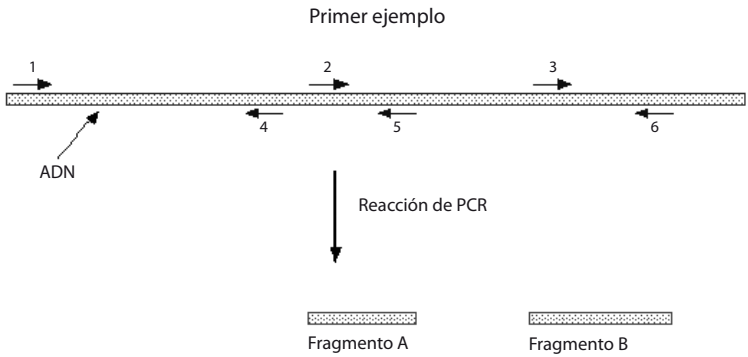
El PCR tiene diferentes métodos o aplicaciones en función de lo que nos interese investigar (como son los RAPDs, AFLPs, ISSRs, SSCP.… véase los capítulos 18 y 19 de este libro) y el primer paso es tener claro el tipo de información que necesitamos para elegir o diseñar la estrategia más apropiada para nuestro trabajo. Brevemente podemos dividir la técnica en dos categorías:

Figura 2. En una reacción de PCR los fragmentos se amplñifican en forma exponencial (tomado de Vierstraete, 2001)

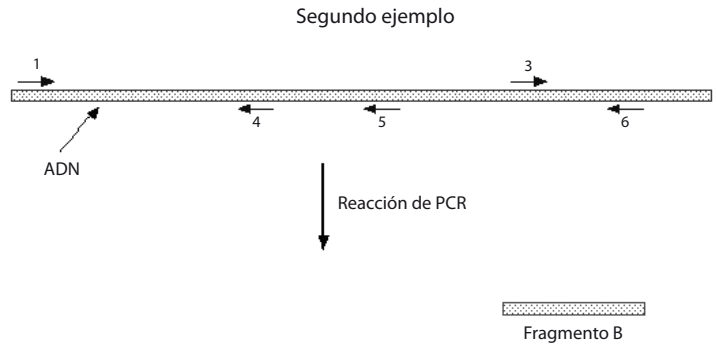


1) PCRs para la amplificación de un solo sitio conocido del genoma (locus). Estos PCRs requieren conocer la secuencia que se trabaja (por ejemplo cuando amplificamos un gen específico como el 16S), en cuyo caso se utilizan oligonucleótidos diseñados a partir de la secuencia de ese gen y se obtiene un solo fragmento de un tamaño ya conocido. Con este tipo de PCRs es posible hacer filogenias, y para obtener los datos hay distintos caminos: desde hacer geles especiales que detectan cambios hasta de una sola base entre las secuencias (SSCP), hasta utilizar enzimas de restricción para generar patrones de cada individuo, aunque lo ideal es obtener la secuencia completa del gen que amplificamos, sobre todo cuando se desea responder a preguntas relacionadas con las fuerzas evolutivas que han actuado sobre él. El gen secuenciado puede ser analizado desde varias perspectivas y muchas de ellas se encuentran en los diversos capítulos de este libro.
2) PCRs en los que no es necesario conocer la región que se está amplificando (se amplifican regiones no conocidas, como zonas hipervariables del genoma), por lo cual no se sabe el tamaño del fragmento (o fragmentos) que se esperan. Éstos se utilizan para determinar polimorfismo genómico y son los más comunes para *fingerprint*, ya que es sencillo obtener los datos (un gel de agarosa después del PCR es suficiente), se observan varios loci simultáneamente y la información de las zonas variables permite inferir los datos necesarios para análisis de genética de poblaciones. En general este tipo de PCRs utiliza un solo oligonucleótido con 2 características importantes: que sea de pequeño a mediano (de 6 a 18 bases) y sobre todo que su secuencia esté presente muchas veces en el ADN del organismo que estudiamos. Existen zonas repetidas hipervariables del ADN que pueden amplificarse de esta manera, por ejemplo los sitios que sirven para iniciar la síntesis de ADN en los cromosomas, conocidas como microsatélites. En el primer ciclo de reacción lo que sucederá es que el oligonucleótido utilizado hibridará en distintas zonas del ADN, y primero comenzarán a sintetizarse fragmentos de tamaños variables e indefinidos (hasta donde la polimerasa logre copiarlos). En el segundo ciclo las cadenas sintetizadas a partir de las primeras copias formadas serán del tamaño que existe entre dos oligonucleótidos que no estén muy alejados entre sí. Estos fragmentos se copiarán una y otra vez, y de esta manera al final obtendremos muchos fragmentos de tamaños diferentes, de los que conoceremos la secuencia con que inician y terminan, pero no la secuencia completa de cada uno. En la figura 3 se explica con más detalle un PCR de este tipo.

Figura 3. Los PCR que amplifican zonas no conocidas utilizan oligonucléotidos pequeños y amplifican zonas repetidas en el genoma (tiomado de Harlocker, 2003)



Supongamos que hacemos una reacción de PCR con un individuo hipotético utilizando un oligo para microsatélites. En este primer ejemplo el oligo ha hibridado en el ADN en 6 sitios diferentes, y al amplificar se generan 2 fragmentos:
1) Fragmento A es sintetizado a partir de la secuencia de ADN que se encuentra entre los oligos que se han unido en las posiciones 2 y 5
2) El fragmento B es sintetizado a partir de los oligos unidos en las posiciones 3 y 6.
No hay productos de PCR entre las posiciones 1 y 4 ya que están muy lejos entre sí como para permitir que ocurra la reacción de PCR. Los oligos unidos a las posiciones 4 y 2 ó 5 y 3 tampoco amplifican ya que no están orientados uno hacia el otro.



En el segundo ejemplo hay que suponer que es un individuo que tiene un cambio en el segundo sitio, y el oligo no se ha unido en la segunda posición, por lo tanto sólo se genera el fragmento B.

¿Qué se necesita para hacer un PCR y cómo conseguir estos reactivos en México?

La polimerasa comercial siempre viene acompañada de un buffer o amortiguador con las sales que se requieren, y si es necesario afinar las condiciones del PCR, se puede pedir el cloruro de magnesio, $MgCl_2$, aparte. Nosotros utilizamos frecuentemente la polimerasa que se produce en la facultad de Veterinaria de la UNAM (*Amplificasa*, de Biotecnologías Universitarias), y nos ha funcionado bastante bien para casi todos nuestros PCRs. Si se trata de un PCR que ha dado problemas, utilizamos una polimerasa más cara (por ejemplo la marca Applied Biosystem). También hemos utilizado distintas marcas de dinucleótidos, y en general todos han funcionado bien. Lo importante es que una vez decididos por una marca, si es posible, hay que tratar de no cambiarla, pues el cambio de marca implica estandarizar otra vez el PCR que ya había salido. Hemos visto que es mejor utilizar la misma marca de polimerasa y de dinucleótidos, ya que cada compañía estandariza sus condiciones con sus propios reactivos, y sobre todo para PCRs difíciles nos ha funcionado mejor de esta manera. Si es un PCR no tan complicado, la polimerasa de la UNAM trabaja bien con distintas marcas de dinucleótidos.

Los oligonucleótidos deben ser sintetizados por alguna compañía comercial, y dentro de la UNAM también hay unidades de síntesis en varios institutos de investigación, como los de Fisiología Celular y Biotecnología. Para mayor información al final del capítulo hay un pequeño directorio con algunos de estos servicios que es posible obtener en la UNAM.

Un asunto crítico que a veces uno no considera es el agua. El agua que se utiliza en una reacción de PCR debe tener muy pocas sales (bidestilada) y si hay variaciones en la cantidad de iones entre una reacción y otra podría haber problemas. Si se cuenta con un buen desionizador en el laboratorio será suficiente con sólo esterilizarla (ver apartado sobre contaminación del PCR para evitar riesgos), si no es así será necesario recurrir al agua que venden las compañías comerciales, embotellada y libre de nucleasas.

Tubos y puntas, será elección de cada quien. Casi todos los termocicladores funcionan con tubos de 0.2 ml cuyas paredes son muy delgadas para que se ajuste mejor la temperatura al interior del tubo cuando se hace la reacción. Hay marcas más baratas, algunas de ellas con la gran desventaja de que las tapas de los tubos no cierran bien y el contenido se evapora, pero hay que probar, pues no siempre es así. Desde luego, todo el material deberá venir certificado: libre de ARNasas y ADNasas.

Los termocicladores también son una elección personal que depende en gran parte del presupuesto. Los hay de todos tipos y para todas las necesidades: que le quepan muchos tubos, que hagan gradientes de temperatura, que sean muy rápidos y exactos para alcanzar las temperaturas programadas… hemos tenido de distintas marcas y nuestra experiencia ha sido que han funcionado mejor los más sencillos que sólo suben y bajan temperaturas con precisión y rapidez. Los termocicladores de modelos más viejitos son menos precisos, ya que los tubos que utilizan son de plástico grueso y la temperatura a la que trabajan no es tan exacta; otra desventaja es que son muy lentos y el PCR tarda mucho en estar listo.

Haciendo PCR: cómo hacerlo y algunos tips para que salga mejor

El PCR es una técnica aparentemente sencilla y fácil de hacer. El problema es que no siempre es así. La reacción de PCR es muy sensible a cambios de iones, temperaturas, contaminantes que pueden estar en el ADN o en el agua… de un termociclador a otro puede haber variaciones ¡y a veces es difícil entender porqué no sale nada! En este apartado trataremos de mencionar todos los detalles que consideramos podrían ser de ayuda, sobre todo para aquéllos que están empezando a montar la técnica.

Antes de empezar

A nosotros nos ha funcionado bien preparar todos los reactivos en alícuotas congeladas (los guardamos a –20ºC), calculando que cada alícuota sirva para unas 5 reacciones como máximo. Sólo para hacer las alícuotas utilizamos puntas especiales con filtro, para evitar contaminarlas, y nos hemos dado cuenta que de esta manera congelamos y descongelamos los reactivos pocas veces, por lo que nos duran más (sobre todo la *taq* y los dNTPs) y además si tenemos algún tipo de contaminación, tiramos las alícuotas que usamos en ese momento, se descongelan nuevas, y de esta forma no tenemos que tirar todos los reactivos a la basura. Es importante considerar que los refrigeradores comerciales con *defrost* automático deshidratan las muestras por lo que hay que utilizar los refrigeradores y congeladores de laboratorio sin esta función.

Condiciones del PCR

Lo mejor es tratar de empezar con las mismas condiciones que se hayan reportado para el oligonucleótido que estemos utilizando. Cuando se trata de un oligo diseñado en el laboratorio (o que no está reportado) se puede empezar a montar la técnica en condiciones estándares, y dependiendo del resultado, se hacen o no modificaciones (tabla 1). Para ello sugerimos probar primero con pocas muestras (por ejemplo, de 5 a 10 ADNs diferentes), y una vez que esté montado el PCR, entonces sí hacer el experimento con las muestras que se necesiten.

Tabla 1. Condiciones estándar para un PCR
(cálculos para 50 µl en cada tubo)

	Concentración inicial*1	Concentración final en la reacción	Cantidad para un tubo	Cantidad para 10 tubos
dNTPs	10mM (todos) *2	200µM varía de acuerdo al magnesio	1 µl	10 µl
Magnesio	25mM	1.5mM puede probarse de 1 a 4 mM	3 µl	30 µl
Oligo forward	10µM	1µM puede probarse de 0.1 a 1 µM	5 µl	50 µl
Oligo reverse	10µM	1µM puede probarse de 0.1 a 1 µM	5 µl	50 µl
Enzima	5U/µl	1U	0.2 µl	2 µl
Buffer	10x	1x	5 µl	50 µl
Agua	--	--	29.8 µl	298 µl
ADN (agregar después de dividir el mix)	0.1mg/ml	0.1mg genómico (máximo 500 ng; de bacteria de 1 a 10 ng y si es plásmido de 0.1 a 1 ng)	11 µl	---

*1Las concentraciones iniciales son las más comunes del mercado, para los dNTPs o los oligos son las concentraciones a las que más comúnmente se preparan.
*2La concentración de los dNTPs a 10 mM, se refiere a la suma de los 4 dNTPs, y cada uno está a 2.5 mM.

En esta tabla hemos anotado las concentraciones finales de cada reactivo, y en las columnas siguientes anotamos cómo haríamos un PCR con estas concentraciones en el laboratorio. Por cada tubo sabemos qué cantidad agregar, y dependiendo del número de muestras que usaremos se calcula la cantidad necesaria de cada reactivo. En este caso pusimos de ejemplo un experimento con 10 muestras: en un tubo eppendorf de 1.5 ml mezclamos las cantidades calculadas (excepto el ADN) y cada uno de los 10 tubos se llena con 49 µl de mezcla, y al final se agrega el ADN de cada muestra.

Una recomendación importante es hacer la mezcla del PCR lo más homogénea posible. Una causa muy común de errores la encontramos aquí, el congelar y descongelar cambia las concentraciones de los reactivos y/o forma gradientes dentro de los tubos, y al homogeneizarlos aseguramos obtener resultados reproducibles. Para hacerlo, nos aseguramos de que todos los reactivos se descongelen totalmente sobre el hielo, y para utilizar siempre la misma cantidad de reactivo en el experimento, cada tubo se invierte y se le dan golpecitos con los dedos; como el contenido queda regado en las paredes utilizamos una picofuga (centrífuga pequeña) para juntar el líquido (también se puede juntar a mano con varias sacudidas fuertes hacia abajo aunque esto no es tan eficiente). Estas indicaciones las seguimos con todos nuestros reactivos, excepto la polimerasa, que nunca sacamos del congelador de –20ºC; cuando necesitamos usarla, en el mismo congelador tomamos lo que se necesite. Al final, cuando ya hicimos la mezcla en el tubo y hemos agregado la *taq*, se invierte suavemente varias veces (sin hacer burbujas pues se desnaturaliza la polimerasa) y se baja la mezcla en la picofuga. Ya preparados, los tubos deben mantenerse en hielo o a 4ºC hasta meterlos al termociclador, para evitar que la polimerasa sintetice fragmentos inespecíficos.

Temperaturas y ciclos

Recomendamos buscar algunos artículos que han trabajado con sistemas parecidos, ya sea el organismo, el gen o el tipo de ADN que se desea trabajar, y empezar a trabajar con las temperaturas que se reportan. Si esto no funciona, o no hay reportes de esos oligonucleótidos, las temperaturas y los ciclos estándares que en general se recomiendan para empezar a probar son los siguientes:

- Desnaturalización inicial: 95ºC por 5 a 10 mn;
- 30 ciclos con desnaturalización a 95ºC por 30 s, alineamiento a 50ºC por 30 s y extensión a 72ºC por tiempo variable (ver más adelante);

- Extensión final a 72ºC por 10 mn (aunque no es estrictamente necesario este último paso, es sólo para asegurar que los fragmentos incompletos se terminen de sintetizar);
- Al final se programa la máquina para que conserve los tubos a 4ºC.

En cuanto al tiempo, existe una regla que puede aplicarse a casi todos los PCRs: la temperatura de desnaturalización y de alineamiento es suficiente de 30 a 60 s. Si es una buena máquina que llega rápidamente a la temperatura programada, con 30 s es suficiente. Para la extensión, dependiendo del tamaño que esperemos se utiliza más o menos tiempo para permitir que la polimerasa trabaje adecuadamente, si se espera un producto de 1 kb, con 1 min es suficiente, si es de 2 kb, 2 minutos, y si es menos, se hace la conversión equivalente (Henegariu, 1997).

Casi todos los PCRs funcionan bien con 30 ciclos, aunque pueden usarse desde 20 ciclos hasta 35. Alguna vez nosotros utilizamos menos ciclos para un PCR que nos daba problemas: por más que hicimos, era imposible quitar una banda inespecífica que se amplificaba en poca cantidad, y bajando los ciclos logramos deshacernos de ella, logrando tener el PCR que esperábamos, de una sola banda (aunque obtuvimos menos producto).

La temperatura de alineamiento

Si con estas temperaturas estándares el PCR no está dando buenos resultados se pueden hacer variaciones en la temperatura de alineamiento. Si la temperatura de alineamiento es muy baja, obtendremos un PCR menos específico, y si es muy alta, la especificidad será mayor (aunque si es demasiado alta no se amplificará nada, pues la unión de los oligonucleótidos con sus sitios complementarios será poco estable y la polimerasa no podrá iniciar la síntesis).

Para PCRs muy específicos en los que se amplifica una sola banda es importante elegir una temperatura de alineamiento que sea la correcta y que nos asegure que el gen que amplificamos es realmente el que queremos. En el caso de PCRs de más de una banda si la temperatura que utilizamos no es tan específica podemos obtener poca reproducibilidad en los experimentos, ya que el oligo se pegará en cualquier parte al azar y no solamente en los sitios que son complementarios, lo que hará que se amplifiquen zonas distintas en un PCR y otro.

Para tener una idea de la temperatura de alineamiento que podría ser la mejor para nuestros oligos podemos calcular la *Tm* de cada oligo que utiliza-

mos. *Tm* significa en inglés *melting temperature* y se refiere a la temperatura a la que se hibridan o se pegan los oligonucleótidos en los sitios que son complementarios. Este proceso dependerá principalmente del tipo de uniones (dobles o triples enlaces de hidrógeno) que formarán sus bases, y por eso la secuencia de cada oligo es la que se toma en cuenta para conocer cuál es la temperatura óptima para su alineamiento. Existen muchas maneras de calcularla, por ejemplo una sencilla es:

Tm = 4(G+C) + 2 (A+T) (Entrala, 2000), aunque hay muchas otras que también pueden utilizarse (hay sitios en internet que lo calculan, ver referencias al final). Conocer la *Tm* puede darnos una idea, pero no siempre es la temperatura que se utiliza en el termociclador, ya que también los iones y otras sustancias que haya en la reacción pueden influir en la forma en que se unen los oligos. Una de estas fórmulas toma en cuenta los iones, pero en general hay que probar experimentalmente hasta encontrar la temperatura óptima.

¿Qué hacer cuando todo empieza a fallar?

Si al probar las condiciones estándares obtenemos demasiadas bandas, o no hay ninguna, algunas posibilidades de lo que esté pasando podrían ser:

1) ¿Cuál es la calidad del ADN que tenemos?

Muchas veces cuando no sale un PCR el problema está aquí (para mayor información puede consultarse el capítulo 16 de este libro). El ADN puede estar contaminado con proteínas o alguna sustancia que inhibe la reacción de PCR. Nuestra experiencia ha sido que ahorra más tiempo y dinero a la larga extraer el ADN con un kit comercial que esté diseñado específicamente para el tipo de organismo con el que trabajemos. En general todas las técnicas de extracción tiene principios básicos comunes: primero romper el tejido y las membranas celulares y nucleares, casi siempre con algún tipo de detergente (por ejemplo el SDS o el CTAB). Si quedan restos de SDS la reacción puede inhibirse, aunque esto puede neutralizarse utilizando 0.5% de Tween 20 o 40 en la reacción de PCR (para más información, consultar el sitio de Lieb B., 2003). Después es necesario deshacerse de todos los componentes celulares que se liberan, y dejar al ADN limpio para poder amplificarlo. A veces se usan sales (como acetatos por ejemplo) que precipitan las proteínas pero no el ADN; o cloroformo y fenol que "atrapan" los lípidos y proteínas y dejan al ADN disuelto en agua. También el cloroformo y el fenol inhiben la reacción de PCR, así como el etanol. Además cada tipo de célula o tejido tiene

componentes característicos, que a veces no podemos quitar y que también inhiben la reacción de PCR, por lo que es necesario investigar qué protocolo es el mejor para el tipo de organismo que trabajemos. Otra posibilidad es que el ADN esté degradado, y para saberlo recomendamos correr una pequeña muestra de nuestro ADN en un gel, hemos visto que de esta manera incluso podemos notar si la cantidad de ADN que estamos utilizando es la correcta o no, y aunque sea indirectamente (no es una medida exacta) podremos hacer un cálculo visual de cuánto ADN tenemos por muestra.

2) Hacer curvas de cloruro de magnesio

Es de las primeras cosas que ajustamos para una reacción de PCR que no ha salido del todo bien. La polimerasa necesita de iones de magnesio para funcionar adecuadamente y en general la concentración de 1.5 mM de magnesio es la más común, pero puede probarse en un rango de 1 a 4 mM. Mucho magnesio inhibe a la polimerasa, y poco puede generar productos inespecíficos.

3) Revisar la concentración y el manejo de los dNTPs

Los dinucleótidos unen iones de magnesio. Para 1.5 mM de magnesio en la reacción de PCR, una concentración ideal de dNTPs es de 200 μM. Pequeños incrementos en la concentración de dNTPs pueden inhibir la reacción porque atraparán el magnesio necesario para que la polimerasa trabaje. Si cambiamos la concentración de dinucleótidos es importante tener en cuenta que existe una relación entre las cantidades de magnesio y las de dinucleótidos en la reacción. También es necesario saber que los dNTPs son muy inestables, si se descongelan más de 3 a 5 veces pueden no funcionar tan bien. Por eso recomendamos hacer pequeñas alícuotas, calculando que sirvan para 3 o 5 experimentos.

4) Revisar el buffer y agregar aditivos

En general cada polimerasa viene con un buffer ya preparado con los reactivos necesarios para que funcione de forma adecuada. Casi todos están preparados con KCl y tris (un buffer 10X estándar contiene 500 mM KCl y 100 mM Tris-HCl, pH 8.3). Algunos buffer incluyen el magnesio necesario para la reacción, y puede suceder que al no notarlo y agregar magnesio extra se esté trabajando en concentraciones mayores a las que suponemos. Será necesario entonces pedir un buffer sin magnesio, y/o tomar en cuenta la concentración del magnesio que incluye (usualmente están preparados para quedar a 1.5 mM

final en la reacción). También es posible utilizar aditivos que en ocasiones mejoran el rendimiento de los PCRs, aunque no siempre funcionan, a veces no tienen ningún efecto o a veces incluso podrían bajar el rendimiento en lugar de aumentarlo, así que hay que probar. Brevemente expondremos aquí algunos de los más comunes (Lieb, 2003). **Formamida**: estabiliza la estructura secundaria del ADN. Generalmente se recomienda utilizarla de 1% hasta 5%. Existen algunos reportes que la utilizan cerca del 10%, pero podría inhibir la reacción de PCR, así que es mejor no utilizar concentraciones mayores a las estrictamente necesarias para una óptima amplificación. **Detergentes no iónicos:** estabilizan la *taq* y evitan la formación de estructuras secundarias. 0.1 a 1% de Tritón X-100, Tween 20 o NP-40 pueden incrementar el rendimiento de la reacción, pero también pueden incrementar la cantidad de productos inespecíficos. **BSA (Bovine Serum Albumin o Albúmina Sérica Bovina):** a concentraciones por encima de 0.8 μg/μl el BSA incrementa la eficiencia de la PCR, ya que actúa como una proteína captadora de iones y otros inhibidores de la *taq* polimerasa como la melanina. **DMSO (Dimetil Sulfóxido):** para algunas amplificaciones podría ser necesario utilizarlo del 2 al 10%, ya que reduce la estructura secundaria del ADN y es útil sobre todo para amplificar regiones con gran cantidad de GCs, aunque se ha visto que al 10% inhibe la actividad de la *taq* en un 50 %.

4) Revisar los oligonucleótidos

Podrían estar degradados, defectuosos, mal diseñados o incluso que la cantidad utilizada no sea la correcta. Si hay mucha cantidad aumenta el rendimiento del PCR, pero podrían formarse productos inespecíficos, y si se agrega aún mayor cantidad los oligos pueden formar dímeros, en vez de unirse al ADN, y eso impide la amplificación del producto. Si utilizamos muy poco podríamos tener mayor especificidad en el producto, pero puede suceder que no veremos el amplificado (pues baja el rendimiento), así que será necesario encontrar el óptimo para nuestra reacción.

Contaminación

Es muy frecuente que los PCRs se contaminen. Para monitorearlo, siempre hay que trabajar con un control negativo: un tubo extra al que se le agreguen todos los reactivos utilizados en el PCR, **excepto ADN**. Si en esta muestra hay amplificación significa que en algún reactivo hay restos de ADN o de producto de PCR que está contaminando el experimento. Si esto sucede, consideramos

que lo más fácil es deshacerse de todas las alícuotas sospechosas (por eso es importante preparar alícuotas de todos los reactivos) y tomar alícuotas nuevas. Para prevenirlo sugerimos algunas medidas que podrían ser de utilidad: los productos de PCR son frecuentemente los que contaminan, al ser moléculas de tamaño muy pequeño y al estar tan concentradas es fácil que una gotita de producto quede en una pipeta (las pipetas absorben el líquido por aspersión y las gotitas quedan regadas como si fueran spray); al preparar una mezcla posteriormente, esto contamina alguno de los reactivos que se utilizan. Para evitarlo existen puntas con filtro y pipetas especiales, pero son caras. También existen cabinas con luz UV y flujo laminar, aunque también son caras. La luz UV forma dímeros de pirimidina entre las dobles cadenas de ADN, por lo que estas moléculas se vuelven imposibles de amplificar. Por otro lado, el flujo laminar evita contaminación de un tubo a otro, y es útil sobre todo si se trabaja con plásmidos o reamplificando productos de PCR ya que las moléculas pequeñas son altamente contaminantes.

Sugerimos también separar las áreas de trabajo (la zona de preparación y la zona de amplificación; si no es posible, limpiar muy bien la zona con alcohol antes de preparar la mezcla del PCR, y trabajar siempre sobre un papel) y tener un juego de pipetas exclusivo para trabajar la mezcla para la reacción de PCR, separadas de las pipetas con las que se manejen los productos de PCR ya amplificados. Otra fuente de contaminación es el ADN que amplificamos (¡ya sea de las muestras o el nuestro!), aquí el problema puede ser sobre todo los dedos, al abrir y cerrar los tubos. La sugerencia más común es utilizar guantes, pero para los que tenemos manos pequeñas pueden ser un gran estorbo y las puntas del guante se atoran en todos lados, por lo que consideramos que debe ser una decisión personal, lo importante entonces es estar consciente de dónde pone uno los dedos. También sugerimos agregar el ADN al final en cada tubo, si es posible utilizar una pipeta sólo para el ADN, y si no hay, tener mucho cuidado y limpiar muy bien las pipetas después de agregarlo. También se evita la contaminación si se guardan en lugares y/o cajas distintas los reactivos del PCR, las muestras de ADN y los productos de PCR.

Otra sugerencia es utilizar material que haya sido esterilizado en autoclave para que esté libre de ADN y de nucleasas. Esta es una medida muy común, pero hemos tenido contaminaciones provenientes del agua de la autoclave, que contenía gran cantidad de algas, y el proceso de esterilización no fue suficiente para eliminar el ADN de estos microorganismos. Desde entonces decidimos no esterilizar el material del PCR, y para descontaminar utilizamos luz ultravioleta en las pipetas; las puntas y tubos los utilizamos directamente

de las bolsas (de marcas certificadas, libres de ADNasas, ARNasas y ADN), con lo cual logramos eliminar la contaminación.

Además del control negativo que nos permite visualizar posibles casos de contaminación, una vez que se ha ajustado el protocolo de nuestro PCR es siempre recomendable incluir un *control positivo*. Éste es un tubo que contiene ADN de una muestra que haya amplificado adecuadamente y cuyo patrón de bandas es conocido. De esta manera, puede esperarse que si la reacción se llevó a cabo correctamente, esta muestra siempre salga. En caso contrario, si nuestro PCR no generó bandas en ninguna de las muestras, ni siquiera en el control positivo, puede suponerse que el error radica en que se nos olvidó agregar algún ingrediente o bien que los ciclos de temperatura no se llevaron a cabo adecuadamente, ya sea por problemas de la máquina o bien por error al elegir el programa de amplificación.

Métodos para visualizar el PCR

Principios básicos de la electroforesis

La idea ahora es poder analizar el o los fragmentos obtenidos en el PCR, y la electroforesis, ya sea en geles de agarosa o de acrilamida, permite separar estos fragmentos de acuerdo al tamaño de cada uno. Tanto la agarosa como la acrilamida forman una especie de red con agujeros de tamaños diferentes, por la cual obligamos a pasar los fragmentos de ADN, "jalándolos" a través de corriente eléctrica, hacia el polo positivo, ya que la carga de una molécula de ADN es negativa por la presencia de grupos fosfato (P^-). Los fragmentos más pequeños pasarán primero a través de la red de agujeros, mientras que los más grandes se irán retrasando y atorando en los hoyos; de esta manera los fragmentos de tamaños similares migrarán a ritmos similares. Si hay muchos fragmentos de un mismo tamaño se agruparán todos juntos, por lo que podremos verlos formando lo que llamamos una **banda** en el gel.

¿Hacemos nuestro gel con agarosa o acrilamida? Las moléculas de acrilamida forman redes con tamaños de poros más uniformes y más pequeños que la agarosa, por lo que este tipo de gel es útil si los tamaños de los fragmentos que manejamos son pequeños; la separación puede ser tan fina que con algunas técnicas es posible separar moléculas de ADN que tienen una sola base de diferencia. El problema es que las técnicas de acrilamida son muy laboriosas y la tinción con plata (si no quiere usarse radioactividad) tiene muchos trucos, por lo que en este apéndice sólo anotaremos algunas recetas para lo más

sencillo, que son los geles de agarosa. Para más información sobre geles de acrilamida pueden consultarse los sitios de internet que sugerimos al final del capítulo. La agarosa no forma redes tan uniformes, pero permite separar las moléculas de ADN en un intervalo muy grande. Utilizarla es muy sencillo y teñir los geles también lo es.

a) Métodos para geles de agarosa

Será necesario tener en el laboratorio equipo que nos permita trabajar con los geles: una cámara de electroforesis, una fuente de poder, un transiluminador de luz UV y equipo de fotografía (lo más sencillo: una cámara polaroid, un filtro para luz UV y un cono adaptado a la cámara, o también existen cámaras especiales y equipo de cómputo específico para ello) para guardar la imagen del gel. Para empezar hay que preparar el buffer de corrida, el cual tendrá el pH requerido y los iones necesarios para que fluya la corriente y pueda migrar el ADN. Si en lugar de buffer utilizáramos agua, el ADN se quedaría inmóvil y no veríamos migración en los geles, y si por el contrario utilizáramos un exceso de sales, el gel se calentaría tanto que al final acabaría por derretirse. El buffer más común es el TBE (Tris Boratos EDTA), que por ser muy estable puede reutilizarse varias veces. También se utiliza con frecuencia el buffer TAE (Tris Acetatos EDTA), que es menos estable que el TBE y tiende a ionizarse más rápido, pero permite obtener mejor separación de bandas, sobre todo si son de gran tamaño (1 kb o más). Las recetas para preparar éstas y otras soluciones están al final de este apartado. Dependiendo del tamaño de los fragmentos que esperamos se utilizará una concentración de agarosa mayor o menor para obtener agujeros menos o más grandes y una mejor resolución de nuestras bandas. Si no conocemos el tamaño, se puede empezar con agarosa al 1%, pero si se conoce, se puede usar esta tabla como guía:

Rango efectivo de separación (kb)	Agarosa (%)
30 a 1	0.5
12 a 0.8	0.7
10 a 0.5	1.0
7 a 0.4	1.2
3 a 0.2	1.5

Fuente: Tomado de Sosnick, 2003.

La agarosa se disuelve en el mismo buffer que utilizaremos para la corrida, y se calienta hasta ebullición para disolver bien el polvo. Hay que tener cuidado

de retirarla del calor o del microondas en cuanto comienza a hervir, pues podría derramarse. Si es un gel muy importante, en el laboratorio pesamos el matraz donde se prepara el gel antes y después de calentarlo, y le agregamos el agua que haya perdido durante el calentamiento, para asegurar que la concentración de agarosa sea la correcta. La solución se agitará suavemente para evitar la formación de burbujas, y cuando se enfríe un poco (aprox. 60ºC) se vierte de una sola vez en el contenedor de geles al que ya le colocamos el peine para que se formen los pozos en donde cargaremos las muestras. Si quedan algunas burbujas, rápidamente con la punta de una pipeta podemos picarlas y quitarlas; si se dejan pueden hacer que la electroforesis no migre en forma homogénea. Cuando se enfríe y solidifique agregamos el buffer necesario hasta cubrir BIEN el gel (hemos visto que de esta forma el peine se puede quitar más fácilmente), y después se retira el peine con cuidado para no romper el fondo de los pozos.

b) Cargando el gel

Con una pipeta, cada muestra se vierte en un pozo, mezclada previamente con 1 ó 2 microlitros de colorante de corrida (receta al final). Generalmente los colorantes de corrida llevan alguna sustancia espesa, como glicerol o sacarosa, que permite que la muestra caiga hacia el fondo del pozo, y los colorantes (como el xilen-cianol o azul de bromofenol) nos dan una idea de cómo van migrando los fragmentos (en un gel de agarosa al 1%, el azul de bromofenol migra junto con los fragmentos de 300 pb, y el xilen-cianol migra igual que los fragmentos de 4 kb). No es necesario utilizar toda la muestra de PCR en una corrida, puede utilizarse del 10% al 20% de la cantidad total del PCR que hicimos (*i.e.* si en total son 50 µl, se cargarán 5 µl de muestra).

Nosotros en un parafilm depositamos unas gotitas de colorante (las necesarias para el número de muestras que usemos) y después agregamos la gotita de la muestra a cargar sobre la gota de colorante. Con cuidado de no hacer burbujas las mezclamos subiendo y bajando con la pipeta. La punta de la pipeta se mete un poco en el pozo (¡sin romperlo!) y lentamente se vacía la pipeta para cargar el gel. Para que las muestras no se derramen y no se mezclen unas con otras hay que evitar llenar el pozo hasta arriba. Por lo menos un carril del gel siempre deberá tener un marcador de peso molecular (muchas casas comerciales los venden) como control para saber el tamaño de las bandas que tendremos. Tampoco debe olvidarse poner en el gel los controles negativo y positivo. El PCR sobrante lo guardamos congelado a -20ºC por si lo necesitamos después. Cuando el gel está listo se conectan los cables, lo más común es un cable rojo para conectarlo al polo positivo y uno negro

en el negativo. El ADN migrará hacia el polo positivo ya que los fosfatos de la molécula le confieren carga negativa, por lo que hay que asegurarse que la corrida del gel sea HACIA el cable rojo, o polo positivo. Para el voltaje, se recomienda utilizar 5 volts por cada centímetro que exista entre los dos electrodos de nuestra cámara (i.e. si la cámara mide 30 cm se correrá a 150V), esta medida se aplica cuando tenemos fragmentos grandes, de más de 2 kb; para un PCR con fragmentos pequeños, de 100 pb hasta 1 kb, nosotros utilizamos casi siempre 90 a 100 volts para geles chicos y/o grandes, y los dejamos 1 ó 2 horas, dependiendo del largo del gel.

c) Tinción del gel

Lo más común es utilizar bromuro de etidio, que es una molécula con dos propiedades importantes: se intercala en las bases del ADN y brilla con luz UV a una longitud de onda determinada (264-366 nm) con lo cual podemos observar las bandas de ADN en el gel. El bromuro de etidio es un mutágeno y es altamente tóxico, por lo cual es necesario utilizar guantes y bata para su manejo. Es recomendable apartar un área del laboratorio y material exclusivo para su uso. Los geles con bromuro deberán juntarse y desecharse con alguna compañía de desechos tóxicos, así como las puntas y guantes contaminados. Las soluciones con bromuro pueden inactivarse, al final hay una lista de 3 sitios en los que dan indicaciones de cómo hacerlo. Si de derrama una pequeña cantidad, hay que absorber muy bien con toallas de papel y después con alcohol (las toallas se juntan en una bolsa para inactivar el bromuro que quede en ellas). Cuando no estamos seguros si algo está o no contaminado con bromuro lo exponemos a la luz UV y si no hay fluorescencia es que no hay bromuro. Hay dos maneras de teñir el gel: cuando la agarosa está a unos 60ºC, antes de vertirla, se añade el bromuro de etidio directamente para que quede a una concentración de 0.5 μg/ml en el gel. De esta forma al terminar la corrida puede verse el gel. La desventaja es que el bromuro retarda la migración de las moléculas de ADN, y además la cámara de electroforesis queda contaminada con el bromuro, es por esto que hay quien prefiere teñirlo después de la corrida: el gel se sumerge en una solución de 0.5 μg/ml de bromuro de etidio por 30 a 40 mn y así no se contaminan las cámaras y el ADN migra más rápido, pero es un método más lento. Para los dos métodos será necesario preparar una reserva de bromuro de etidio como se indica al final del capítulo.Al terminar se pone el gel en el transiluminador para verlo. La luz UV puede dañar la piel y los ojos, por lo que será necesario protegerse a través de un acrílico especial y protectores de la cara y/o los ojos.

Recetas

EDTA	Para preparar ambos buffer TBE y TAE es necesario tener una solución de EDTA 0.5M pH8. Para preparar 100 ml, pesar 18.61g de EDTA y disolver en 70 ml agua. Sólo se disolverá hasta que la solución alcance un pH de 8. Ajustar primero con lunetas de sosa (unas 10), dejar que se disuelvan e ir agregando poco a poco más lunetas hasta llegar a pH de 6 o 7, y aquí ajustar hasta 8 con gotitas de sosa 5 o 10 M, después aforar a 100 ml.
TBE	Se prepara un stock concentrado a 5X, y para hacer el gel, se diluye y se utiliza a 1X o 0.5X. Es común que el TBE se precipite, por eso hay que preparar cantidades que se utilicen constantemente. La receta del TBE 1X es: 0.089M tris, 0.089M ácido bórico y 0.002M EDTA. Para preparar una solución 5X, pesar 54 g de tris-base y 27.5 g de ácido bórico. Disolver en 800 ml de agua y añadir 20 ml de EDTA 0.5M pH8. Aforar a 1 litro. Una solución de TBE bien preparada deberá tener un pH cercano a 8.3 sin necesidad de ajustarla.
TAE	No se precipita y es posible prepararlo muy concentrado, aquí escribiremos la receta al 50X y para la corrida se diluye al 1X. TAE 1X: 0.04M tris-acetatos, 0.002M EDTA. Para una solución 50X: pesar 242 g de tris base y disolverlo en 500 ml de agua estéril, agregar 57.1 ml de ácido acético glacial y 100 ml de EDTA 0.5 M pH 8. Aforar a 1 litro.
Colorante de corrida	xilen–cianol al 0.02%, azul de bromofenol al 0.02 % en glicerol al 50%.
Bromuro de etidio	La costumbre es hacerlo a 10 mg/ml, se guarda a 4ºC protegido de la luz en una botella oscura, envuelta con aluminio.

Bibliografía

Información general de la técnica

Afseth G.1997. PCR primer, stategies to improve results http://www.biotechlab.nwu.edu/pe/ , accesado 08/03.

Casas Ciria F.J, 2003. Microbiología química en la www http://www.microbiologia-clinica.com/generalidades.htm, accesado 08/01.

Entrala C. 2000. Técnicas de análisis del adn en genética forense http://www.ugr.es/~eianez/Biotecnologia/forensetec.htm, accesado 08/01.

Prilusky J. 2003. BioGuide–PCR http://bioinformatics.weizmann.ac.il/mb/bioguide/pcr/contents.html, accesado 08/01.
Roolpi P. 2002. Polymerase Chain Reaction http://palou.uib.es/roolpi/PCR/ accesado 08/03.
Vierstraete A. 2001. Principle of the PCR http://allserv.rug.ac.be/~avierstr/principles/pcr.html, accesado 08/01.

Sitios comerciales

Biotools. 2003. Amplificación de ADN http://www.biotools.net/esp/tecnica/t3.htm, accesado 08/03.
Roche. 2000. Innovative tools for amplification http://www.roche-applied-science.com/pcr/application_hints_01_4a1.htm, accesado 08/03.

Sitios con tips cuando hay problemas

Frame P. 2002. Ten Things That Can Kill Your PCR http://www.biowire.com/nucleus/nucleus_1_1.jsp accesado 08/03.
Henegariu O. 1997. PCR and multiplex PCR: guide and troubleshooting http://info.med.yale.edu/genetics/ward/tavi/PCR.html accesado 08/03.
Krivokapich S. J. 2003. PCR links.com la guía web de la técnica de PCR http://www.espanol.pcrlinks.com/generalidades/aditivos.htm, accesado 08/03.
Lieb B. 2003. PCR additives http://www.uni-mainz.de/~lieb/additiva.html, accesado 08/03.

Sitios con tipos de PCRs (RAPDS, AFLPs, ISSRs...)

Gillet E. (2000) Which marker for wich purpuose? http://www.metla.fi/archive/forestgen/2000/02/msg00010.html, accesado 08/03.
Harlocker S (2003) RAPD PCR http://avery.rutgers.edu/WSSP/StudentScholars/project/archives/onions/rapd.html , accesado 08/03.
Howis S.(2001) PCR techniques http://rucus.ru.ac.za/~wolfman/Essays/PCR1.html accesado 08/03.
SIUC / College of Science (2001) DNA-Based techniques for studies at population & species level. http://www.science.siu.edu/Plant-Biology/PLB449/DLN.449lectures/PCR&RAPD.htm, accesado 06/03.

Sitios con información sobre geles de agarosa o acrilamida

Binder A. 1997. Agarose gel electrophoresis http://www-ang.kfunigraz.ac.at/~binder/thesis/node41.html, accesado 08/03.
Wallman S. 2001. Technical tutorials: Polyacrilamide gel electrophoresis http://biotech.tec.nh.us/BT210/tutorial10.html, accesado 08/03.
Sosnick T. 2003. Agarose gels http://sosnick.uchicago.edu/agarose_gels.html.

Sitios para descontaminación de Bromuro de Etidio

OEHS. 1998. Methods for the destruction and decontamination of ethidium bromide http://www.utexas.edu/safety/ehs/resources/info.ethidium.pdf accesado 08/03.
ORS Chemical Safety. 2001. Decontamination of Ethidium Bromide http://www.northwestern.edu/research-safety/chem/ethid2.htm, accesado 08/03.
UCSC Laboratory Safety Services. 1990. Ethidium bromide, chemical information sheet http://ehs.ucsc.edu/Lab_Research_Safety/Pubs/Facts/ETBR.pdf, accesado 08/03.

Sitios con otras referencias y/o vínculos

Gannon P. 2000. Cell and molecular biology on line http://www.cellbio.com/protocols.html, accesado 08/03.
Highveld. 2003. The World Famous PCR Jump Station http://highveld.com/f/fpcr.html, accesado 08/03.

Calcular Tm

Buehler E. 2003. Oligo calculador http://www.pitt.edu/~rsup/OligoCalc.html. Consultado el 08/03.
Cao Q. 2002. Oligonucleotide Properties Calculator http://www.basic.nwu.edu/biotools/oligocalc.html , accesado 08/03.
Proligo. 1998. PROLIGO - Oligos Parameter Calculation http://www.gensetoligos.com/Calculation/calculation.html , accesado 08/03.

Bibliografía sugerida que puede conseguirse en la UNAM

Bej, A.K., M.H. Mahbubani y R.M. Atlas. 1991. Amplification of nucleic acids by polymerase chain reaction (PCR) and other methods and their applications. *Critical*

Reviews in Biochemistry and Molecular Biology 26:301-334.
Burke J. (ed.). 1996. PCR : essential techniques. Wiley Europe, New York, EUA.
Dieffenbach, C.W y G.S. Dvksler. 1995. PCR primer: a laboratory manual. CSHL Press, Cold Spring Harbor, EUA.
Eeles, Rosalind A.. 1993. Polymerase chain reaction (PCR) : The technique and its appplication. R. G. Landes, Austin.
Erlich, H. A. 1989. PCR technology: Principles and applications for DNA amplification. Stockton Press, New York, EUA.
Innis M.A., H. David, D.H. Gelfand y J.J. Sninsky (eds.). 1995. PCR strategies. Academic Press, San Diego, EUA.
Mcpherson M.J., P. Quirke y G.R. Taylor. 1995. Polymerase chain reaction : a practical approach . Oxford university, New York, EUA.
McPherson M.J., B.D. Hames y G.R. Taylor. 1995. PCR 2 a practical approach . Oxford University, New York, EUA.
Mullis K.B., Ferre F., Gibbs R.A eds. 1994. The polymerase chain reaction. Birkhauser, Boston, EUA.
Newton C.R. ed. 1995. PCR essential data. Wiley Europe. Chichester USA.
Rapley R ed. 1996. PCR sequencing protocols. Methods in molecular biology: 65. Humana Press, Totowa, New Jersey.
White B.A. (ed.). 1993. PCR protocols: current methods and applications. Humana Press, Totowa, New Jersey.
White B. A. (ed.). 1997. PCR cloning protocols : from molecular cloning to genetic engineering. Methods in molecular biology: 67. Humana Press. Totowa, New Jersey.

	Número de referencia												
Lugar en la UNAM	1	2	3	4	5	6	7	8	9	10	11	12	13
Centro de Neurobiología				x		x	x	x					
Centro Inv. Fijación Nitróg.								x			x	x	
ENEP iztacala			x		x		x						
Facultad de Ciencias		x			x				x		x	x	x
Facultad de Medicina			x			x							
Facultad de Medicina. Investigaciones Clínicas			x		x		x					x	
Facultad de Med. Veterinaria							x	x					
Facultad de Química			x				x	x	x				
Facultad de Química, Posg.			x				x					x	
FES Cuahutitlán							x						

(Continúa)

	Número de referencia												
Lugar en la UNAM	1	2	3	4	5	6	7	8	9	10	11	12	13
FES Zaragoza Campus 1							x						
FES Zaragoza Campus 2				x			x	x			x	x	x
Instituto de Biología		x									x		
Instituto de Ecología			x		x								
Instituto de Fisiología celular			x	x	x	x		x	x	x		x	x
Inst. Inv. Biomédicas	x		x	x	x	x	x	x			x	x	x
Instituto de Química												x	x
Programa Universitario del medio ambiente									x				
UACP CCH			x		x		x						

Direcciones para conseguir oligos o TAQ en la UNAM

1. Unidad de Síntesis. Instituto de Biotecnología, UNAM. Apdo. Postal 510-3 Cuernavaca, Morelos, MEXICO. C.P. 62250. Tel. (01777) 329 1604, (0155) 5622 7604; Fax.(01777) 317 2388. Responsable: Dr. Rubén Paul Gaytán Colín. http://www.ibt.unam.mx/sintesis/oligos.html.
2. Unidad de Síntesis. Instituto de Fisiología Celular, UNAM. Circuito Exterior, Ciudad Universitaria, 04510, México DF. Responsable: Dra Laura Ongay. Teléfono 56 22 56 52. http://www.ifisiol.unam.mx/UBM.html
3. Facultad de Medicina Veterinaria. Responsable: Dr. Rogelio Alonso. Tel 56 22 58 94.

Capítulo 18

Breve revisión de los marcadores moleculares

Miroslava Rentaría Alcántara

Variación morfológica y variación molecular

Hoy en día los biólogos evolutivos usan tanto datos morfológicos como moleculares para establecer hipótesis de relaciones filogenéticas entre organismos, para estimar la variación dentro de las poblaciones y para probar hipótesis de adaptaciones ecológicas. Sin embargo, es común observar incongruencias entre los análisis basados en datos morfológicos y los basados en datos moleculares (Hillis y Wiens, 2000), lo que ha originado polémicas respecto a qué tipo de datos pueden proveer de información adecuada para sustentar y probar hipótesis evolutivas.

El principal argumento en favor de la utilización de caracteres moleculares es que son universales. En muchos casos, principalmente cuando se requiere comparar linajes con divergencia temprana, es imposible establecer hipótesis de homología morfológica; en cambio, existen genes presentes en todos los genomas celulares –como los ribosomales–, que pueden proveer de información para reconstrucciones filogenéticas, donde los caracteres morfológicos son inaplicables (Avise, 1994). Además, estudios teóricos y empíricos han mostrado que en la reconstrucción de filogenias es crucial contar con numerosos caracteres informativos (Hillis *et al*., 1994) y los estudios típicos de secuencias nucleotídicas implican varios cientos o miles de caracteres en contraste con los estudios morfológicos, en los que un análisis incluye

raramente más de cien caracteres (Sanderson y Donoghue, 1989). Los datos moleculares también tienen la ventaja de trabajar directamente con la base genética de la variación, mientras que la base genética de la mayoría de los caracteres morfológicos se asume. Asimismo, en el acercamiento molecular los caracteres se pueden seleccionar y definir de una manera relativamente objetiva (Hillis y Wiens, 2000).

En los estudios morfológicos, en cambio, los caracteres deben ser descubiertos y delimitados generalmente sin ningún criterio explícito para la selección o la codificación del carácter, por lo que tienen el potencial de ser arbitrarios. Por ejemplo, los morfologistas no divulgan generalmente sus criterios para incluir o excluir caracteres y cuando se dan los criterios, varían considerablemente entre estudios (Hillis y Wiens, 2000).

Sin embargo, tienen la ventaja de permitir un muestreo taxonómico mucho más cuidadoso que el que se realiza con análisis moleculares, lo que es importante para las revisiones sistemáticas, los estudios de la evolución del carácter y la valoración filogenética. Asimismo, los especímenes de museo ofrecen muchos caracteres morfológicos para una gran cantidad de taxa, mientras que un muestreo de este tipo para un estudio molecular puede ser difícil por el alto costo de la secuenciación, la necesidad de material relativamente fresco y la inaccesibilidad de las áreas donde se distribuyen algunos de ellos (Hillis y Wiens, 2000).

La morfología es también la única manera en que la mayoría de los taxa fósiles pueden analizarse filogenéticamente y las especies extintas no sólo representan una gran proporción de la biodiversidad de la tierra, sino que también pueden ser cruciales para el entendimiento de las relaciones entre los taxas vivos (Smith, 1998).

De lo anterior se puede apreciar que tanto los acercamientos moleculares como morfológicos tienen ventajas y desventajas; que ambos enfoques siguen desempeñando un papel crucial en casi todos los grupos de organismos y que hasta nuestros días las especies se describen y se identifican con base en ambas clases de datos.

Marcadores moleculares

Los marcadores moleculares son una herramienta necesaria en muchos campos de la biología como evolución, ecología, bio-medicina, ciencias forenses y estudios de diversidad. Además se utilizan para localizar y aislar genes de interés. En la actualidad existen varias técnicas moleculares que

nos permiten conocer cómo se encuentran las proporciones de genes en las poblaciones naturales de manera indirecta, como con los análisis de proteínas, o de manera directa con estudios de ADN. Los diferentes tipos de marcadores se distinguen por su capacidad de detectar polimorfismos en loci únicos o múltiples y son de tipo dominante o co-dominante (Simpson, 1997).

ADN nuclear (nADN)

En eucariontes, la mayor parte de la información genética se encuentra contenida en el núcleo de la célula. El nADN se encuentra empaquetado y asociado a proteínas histonas, conformando los cromosomas.

El nADN contiene regiones únicas –de una sola copia– y no únicas –duplicadas o regiones repetitivas (Stansfield, 1992). Se considera que los organismos diploides tienen dos copias de cada región genética (locus) en los pares homólogos de los cromosomas, llamadas alelos, sin tener en cuenta si contienen regiones codificantes (exones) o no codificantes (intrones o regiones intergénicas).

ADN de cloroplasto (clADN)

En los organismos fotosintéticos con cloroplastos existe un ADN típicamente bacteriano circular, de 120 a 200 kb (Brown *et al*. 1979), con intrones y exones que se considera muy conservado, ya que se trata fundamentalmente del mismo genoma desde las hepáticas hasta las plantas superiores. Cada cloroplasto contiene varias regiones nucleotídicas, cada una con 8 a 10 moléculas de ADN. Un organismo unicelular como *Euglena* puede contener de 40 a 50 cloroplastos, por lo que la celula entera puede contener más de 500 copias del genoma del cloroplasto (Stansfield, 1992).

ADN mitocondrial (mtADN)

El genoma mitocondrial (mtADN) tiene un tamaño de 15 a 17 kb (Brown *et al*. 1979) y su longitud varía considerablemente entre especies: 20 micrómetros en *Neurospora*; 25 micrómetros en levaduras; 30 micrómetros en plantas superiores; 5 micrómetros en algunos animales metazoarios (multicelulares). Se considera que la mayoría del mtADN de hongos y plantas no codifica, ya que el mtARN contiene intrones. El mtARN de animales carece de intrones y

se transcribe como ARN policistrónico que se parte en ARN monocistrónico antes de la traducción (Stansfield, 1992).

Las moléculas de cloroplasto y mitocondria son especialmente importantes para trazar historias filogeográficas y de estructura poblacional genética estrechamente relacionada al linaje, porque son de herencia uniparental y no recombinan. También nos permiten inferir cambios demográficos y de dispersión entre especies (Dirienzo y Wilson, 1991).

ADN ribosomal (RADN)

El rADN puede encontrarse en mitocondrias, cloroplasto y núcleo. Contiene la información para el ARN que conforma los ribosomas, por lo que es información que se transcribe pero no se traduce. El rADN se presenta en repeticiones tándem y está formado por tres subunidades altamente conservadas (18 rADN, 5.8 rADN y 28 rADN), separadas por dos espaciadores con elevadas tasas de sustitución (ITS1 e ITS2). Estas repeticiones en tandém se encuentran conservadas a lo largo de todo un genoma y evolucionan concertadamente, lo que se atribuye a eventos recombinatorios como entrecruzamiento desigual y conversión génica. Estas secuencias, por la baja tasa de sustitución que presentan, son extremadamente útiles en el planteamiento de hipótesis de relaciones filogenéticas de taxa con tiempos de divergencia muy antiguos (Hills *et al*., 1991).

Isoenzimas/aloenzimas

Las isoenzimas fueron originalmente definidas como formas moleculares múltiples de las enzimas que tienen funciones idénticas o similares y están presentes en el mismo individuo (Market y Moller, 1959).

Las isoenzimas pueden presentar varias formas alélicas, conocidas como aloenzimas (Prakash *et al*., 1969). En su mayoría son selectivamente neutras y se utilizan como marcadores hereditarios para cuantificar las frecuencias alélicas y genotípicas de los individuos, que son los estimadores básicos de la composición genética de una población. En estudios de genética las isoenzimas fueron usadas desde 1966, cuantificando la variación genética en poblaciones humanas y de *Drosophila* y un poco más tarde en estudios de plantas superiores (Harris, 1966; Johnson *et al*., 1966; Hubby y Lewontin, 1966; figura 1).

La aplicación de las isoenzimas está dirigida a la cuantificación de heterocigosis, diversidad genética, diferenciación genética y otras medidas de variación genética intra e interpoblacional. También han sido aplicadas exitosamente

para evaluar y entender aspectos de biología evolutiva como los sistemas de reproducción y patrones de fecundación cruzada, relaciones entre fenotipo y ambiente, filogenias, endemismo, diversidad en plantas clonales y apomícticas e interacciones planta-animal (Pérez-Nasser y Piñero, 1997) (véanse los capítulos 6 y 7 de este libro).

Por ejemplo, en el estudio de los sistemas reproductivos de plantas las aloenzimas tienen las siguientes ventajas: 1) son expresadas codominantemente, por lo que los genotipos homócigos y heterócigos pueden ser distinguidos con mucha precisión, y es poco probable que afecten el comportamiento del polinizador; 2) muchos de sus loci son polimórficos, casi cualquier especie presenta al menos uno o dos; 3) probablemente no están sujetas a fuertes fuerzas selectivas, por lo que se consideran buenos marcadores (Brown *et al*. 1989); 4) la técnica es barata en comparación con otras.

Por otro lado, presentan algunas limitaciones: revelan poca variación y la técnica es muy laboriosa, requiere de mucho tiempo y entrenamiento y, sobre todo, es poco reproducible entre laboratorios.

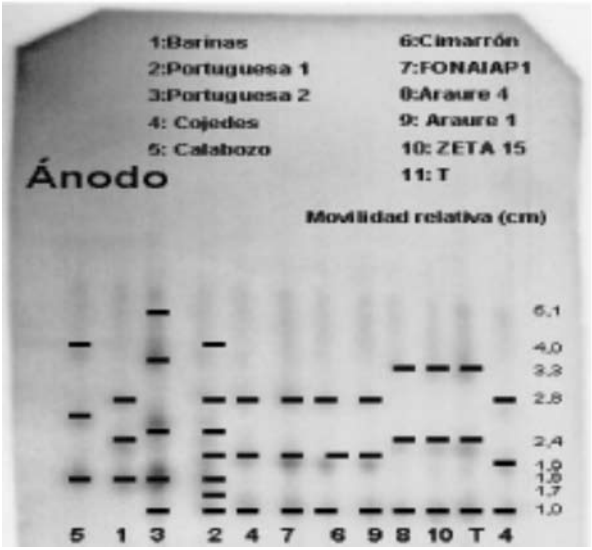
En la electroforesis de enzimas se utilizan cuatro métodos principales: a) gel de almidón (horizontal y vertical); b) gel de poliacrilamida; c) gel de agarosa y d) gel de acetato de celulosa. La electroforesis horizontal en geles de almidón sigue siendo la preferida, ya que es relativamente barata, de fácil manipulación, buena resolución, no es tóxica y se puede analizar de forma rápida la variación para uno o varios loci.

Para un locus dado, el número de bandas varía entre individuos homócigos y heterócigos (figura 1) y también varía en función de la estructura cuaternaria de la enzima; si es monomérica los individuos homócigos representan una sola banda y los heterócigos presentan dos bandas; si es una enzima dimérica, los heterócigos presentan tres bandas (dos homodímeros de los padres y un producto adicional de movilidad intermedia); en las enzimas tetraméricas se observan 5 bandas en los individuos heterócigos.

RAPDs

Los RAPDs (polimorfismos de ADN amplificados al azar) son marcadores que amplifican aleatoriamente segmentos de ADN en una gran variedad de especies. Los RAPDs se basan en la probabilidad estadística de que se presenten sitios complementarios al oligonucleótido de 10 pares de bases (pb) a lo largo del genoma. El polimorfismo de las bandas entre los individuos se debe a cambios en la secuencia de los nucleótidos en los sitios de acoplamiento del

Figura 1. Patrones de la fosfoglucosa isomerasa para ecotipos de arroz rojo y variedades de arroz en gel de almidón de papa agarosa al 12% (foto de Ortiz *et al.*, 2002)



oligonucleótido y por inserción o deleción de los fragmentos en estos sitios (Williams *et al.*, 1990; figura 2).

Estos marcadores son dominantes, es decir, no pueden discernir los homócigos dominantes de los heterócigos para un segmento particular (Whitkus *et al.*, 1994; Backeljau *et al.*, 1995), por lo que la estimación de las frecuencias alélicas se debe hacer de manera indirecta, asumiendo equilibrio de Hardy-Weinberg (Aagaard *et al.*, 1998).

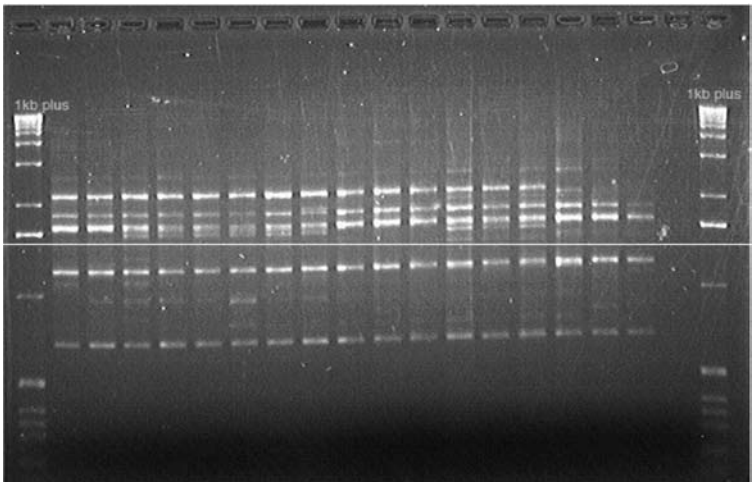
Los RAPDs son útiles en la elaboración de mapas genéticos, en el estudio de parentesco y en el análisis de la estructura poblacional, ya que ayudan a estimar tamaño efectivo, aislamiento reproductivo y niveles de fecundación cruzada (Otero *et al.*, 1997; Parker *et al.*, 1998; véanse también los capítulos 2, 3 y 6 de este libro). Dado el gran polimorfismo que detectan una de sus mejores aplicaciones es la identificación genética de individuos, que incluye casos de clones, híbridos somáticos y mutantes. Otra aplicación paralela es la detección de uniformidad genética con un marcador eficiente y rápido, lo cual puede ser útil en la determinación de estabilidad en programas de reforestación (Otero *et al.*, 1997).

Entre las principales ventajas de los RAPDs esta que amplifican regiones tanto codificantes del ADN como las no codificantes y revelan niveles de variación más altos que los RFLPs e isoenzimas (Williams *et al*., 1990; Lynch y Milligan, 1994; Otero *et al*., 1997; Russell *et al*., 1997; Parker *et al*., 1998); es una técnica relativamente fácil que no necesita conocimiento previo de la secuencia de ADN, no requiere la construcción o el mantenimiento de una librería genómica, el número de loci que puede ser examinado es ilimitado y no requiere pruebas radiactivas (Reiter *et al*., 1992; Whitkus *et al*. 1994).

Sin embargo los problemas prácticos detectados con los RAPDs son la presencia de bandas "erróneas" (artefactos), la reproducibilidad de los resultados y la comigración de bandas. Asimismo, muchos de los alelos raros presentes en las poblaciones estudiadas con RAPDs no son detectados o pueden ser mal interpretados (Zhivotovsky, 1999). Por otro lado, como los loci son dominantes, los RAPDs dan menos información genética por locus que los marcadores codominantes.

Al analizar los datos obtenidos con RAPDs se deben tener en cuenta dos supuestos: 1) cada uno de los marcadores representa un locus mendeliano en el cual el marcador visible, el alelo dominante, está en equilibrio de Hardy-Weinberg con un alelo recesivo y 2) los alelos marcados para diferentes loci no migran a la misma posición en el gel (Lynch y Milligan, 1994).

Figura 2. Patrones de bandeo con RAPDs en *Ferocactus robustus* en gel de agarosa y tinción con bromuro de etidio (foto de Israel Carrillo).



Microsatélites

Los microsatélites o secuencias simples repetidas (SSRs; Litt y Luty, 1989) son secuencias de ADN formadas de 1 a 4 pares de bases, por ejemplo mononucleótidos $(TT)_n$, dinucleótidos $(AT)_n$, o tetranucleótidos $(AAGG)_n$. Estos loci se encuentran en regiones codificantes y no codificantes del ADN y es probable que se formen por eventos de rompimiento que generan polimorfismos con valores superiores al 90% (Armour *et al*., 1994; Coltman *et al*., 1996; Gupta *et al*., 1996; figura 3).

Los microsatélites de ADN nuclear han sido detectados en múltiples grupos de plantas y animales, y han sido utilizados fundamentalmente para estudios de variación genética intra e interespecífica (Edwars *et al*., 1991; Armour *et al*., 1993; Devey *et al*., 1996; Queller *et al*., 1993; Weight y Bentzen, 1994), análisis de linajes (Queller *et al*., 1993) y de sistemas reproductivos (Awadalla y Ritland, 1997). Recientemente se han encontrado microsatélites en algunos organelos citoplasmáticos, como el cloroplasto (SSRc; Powell *et al*., 1995a, b; Vendramín *et al*., 1996; figura 3) y la mitocondria (SSRm; Soranzo *et al*., 1998) lo que ha enriquecido la fuente de estudios evolutivos, ya que estos organelos son heredados uniparentalmente y no están sujetos a recombinación, por lo que los cambios acumulados que observamos en las poblaciones se deben sólo a los procesos de mutación y demográficos (Echt *et al*., 1998). Esto permite contestar preguntas evolutivas muy puntuales relacionadas con el monitoreo del flujo genético, introgresión (Vendramin *et al*., 1998), análisis de paternidad (McCracken *et al*., 1999), para hacer inferencias de parámetros demográficos y para determinar patrones evolutivos de los procesos históricos del origen de especies, formas o razas (Golstein *et al*., 1996).

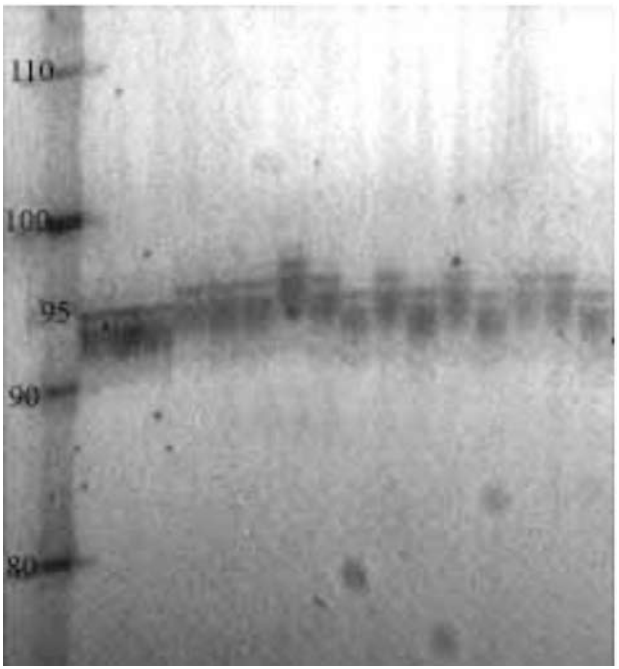
Los microsatélites han tomado ventaja sobre otros marcadores genéticos como los AFLPs, RAPDs, RFLPs, debido a que: i) tienen el más alto grado de polimorfismo; ii) segregan de manera mendeliana y son codominantes; iii) la presencia de un solo locus genético por microsatélite hace que la lectura de las bandas sea clara y fácil de interpretar y iv) son selectivamente neutros (Golstein y Pollock, 1994; Vendramin *et al*., 1996). Además, para trabajar con SRR es necesario conocer la secuencia de la región a analizar para contar con primers específicos que amplifiquen la región repetitiva (el microsatélite) responsable de la variación observada, que además es homóloga para diferentes especies o incluso géneros (Golstein *et al*., 1996). Esto es, los microsatélites son específicos para ciertos grupos de especies y

homólogos entre sí (Vendramin *et al*., 1996), lo que permite hacer estudios comparativos entre especies o géneros de un mismo grupo. Por otra parte, se han realizado estimaciones de las tasas de mutación de estas regiones y se ha llegado a la conclusión de que los microsatélite del ADN nuclear tienen tasas de mutación de $1x\ 10^{-3}$ a $1x\ 10^{-6}$ (Wiessenbach *et al*., 1992; Weber y Wong, 1993), más altas que las que se presentan en el ADN de cloroplasto las cuales según Provan *et al*. (1999) son de $3.2\text{-}7.9 \times 10^{-5}$. Conocer las tasas de mutación nos brinda una base importante para realizar análisis robustos de la genealogía de las poblaciones que nos hablen acerca de la historia evolutiva de las especies, ventaja muy importante sobre otros marcadores genéticos.

Otra ventaja que presenta el uso de microsatélites con relación al uso de secuencias, es que la mutación es homogénea: en la mayoría de los SSR las mutaciones son de un paso (una unidad repetitiva) siguiendo el modelo mutacional SSM de Ohta y Kimura (1973). Sin embargo el número de mutaciones varía si se trata de diferentes tipos de microsatélites, por ejemplo, Golstein *et al*. (1995b) mencionan que las mutaciones de repeticiones formadas por dinucleótidos o trinucleótidos son de dos pasos o incluso de múltiples pasos, lo que ha sido confirmado por estudios en trinucleótidos. Por ejemplo, en el humano el microsatélite formado por las repeticiones CAG – asociado con un desorden neurológico–, puede mutar de pocos a cientos de alelos (Ashley y Warren 1995). Sin embargo este comportamiento asimétrico del tamaño de las mutaciones de SSR ha sido raramente reportado.

Al ser los microsatélites altamente polimórficos, pueden ser ventajosos con relación al uso de secuencias, aunque la variación del ADN nuclear debe analizarse con precaución, ya que pudo haberse generado por duplicaciones o por la presencia de familias multigénicas (Rieseberg,1991), que generan un alto grado de homogeneidad dentro y entre especies, proceso conocido como evolución concertada. Estos fenómenos pueden distorsionar la historia evolutiva de los organismos en estudio y confundir las relaciones filogenéticas (Rieseberg *et al*., 1991b). Es decir, se pueden presentar aparentes reticulaciones o escenarios de hibridización introgresiva. Finalmente, los microsatélites han sido utilizados para hacer reconstrucciones de árboles de genes con base en la teoría de coalescencia (véase el capítulo 4 de este libro). Los fragmentos son separados en geles de acrilamida y son visualizados con nitrato de plata o radioactividad.

Figura 3. Patrones del microsatélite Pt63718 de cloroplasto en *Pinus strobiformis* en gel de acrilamida al 6% y tinción con nitrato de plata (foto de Alejandra Moreno Letelier)



ISSRs

Es un marcador molecular conocido como secuencias repetidas intersimples. Esta es una técnica relativamente nueva y es similar a los RAPDs, excepto que en los ISSRs el primer es un di ó trinucleotido repetido (Culler y Wolfe, 2001 véase el capítulo 19 de este libro).

Los dos métodos que se utilizan en la electroforesis de ISSRs son: a) gel de agarosa visualizado con bromuro de etidio, y b) geles de acrilamida teñidos con nitrato de plata.

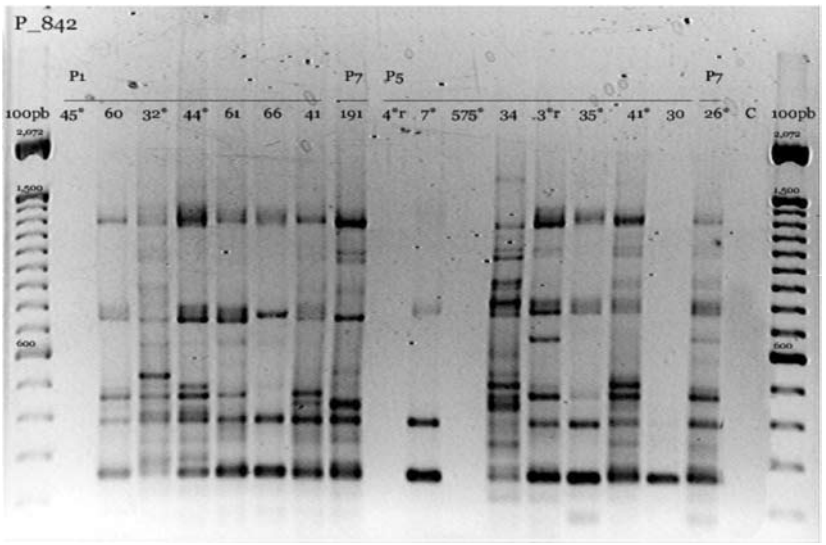
Las bandas que se obtienen con este marcador van de 100 a 2000 pb. La variación alélica en los ISSRs consiste de la presencia o ausencia de los productos amplificados.

Comparada con los primers de los RAPDs las secuencias de los ISSRs es usualmente larga lo que resulta en una mayor reproductibilidad de las bandas.

Los ISSRs han sido utilizado en especies cultivadas desde 1994, pero sólo recientemente se han utilizado en estudios de la variación poblacional en especies silvestres (Wolfe y Liston, 1998; Wolfe et al. 1998 a,b). Hasta la fecha sólo existen algunos estudios donde se ha comparado directamente la variación genética, incluyendo estimaciones de la tasa de fecundación cruzada (que tradicionalmente se estima con marcadores codominantes) y se ha utilizado esta técnica en especies donde la variación poblacional con enzimas es pequeña o no existe. Dada la alta variación genética entre individuos de una misma población, sería recomendable utilizar estos marcadores para análisis de paternidad.

Entre las ventajas principales de los ISSRs es que tienen un gran número de bandas polimorficas. En un estudio realizado en *Viola pubescens* una planta cleistogamica se reportó una gran variación genética. A nivel de especie, el 100% de los loci fueron polimorficos aún cuando los *primers* (se utilizaron tres) fueron seleccionados al azar. Dentro de cada población cerca del 71% de 83 loci fueron polimorficos (Culley y Wolfe 2001). Además es una técnica relativamente fácil de montar, altamente repetible y ya existen primers universales para plantas.

Figura 4. Patrones de ISSRs en *Opuntia rastrera* en gel de agarosa al 1.5% y tinción con bromuro de etidio (foto de Lucia Plasencia)



Por otro lado las limitaciones de los ISSRs son: que las bandas son leídas como marcadores dominantes, es decir cuando tenemos la presencia de la banda no se sabe si el individuo es homócigo dominante o herócigo, que la diversidad genética está basada considerando que cada banda representa un locus con dos alelos y que el alelo dominante esta en equilibrio de Hardy-Weinberg con un alelo recesivo y por último que es una técnica relativamente cara, ya que involucra el uso de geles de poliacrilamida y para su visualización nitrato de plata.

RFLPs

El análisis de polimorfismos de la longitud de los fragmentos de restricción (RFLPs) fue el primer marcador de ADN utilizado por biólogos poblacionales (Parker *et al*., 1998). Este método expresa diferencias específicas del ADN que fueron reconocidas por enzimas de restricción particulares (endonucleasas). Cada una de las endonucleasas (de origen bacteriano), reconoce y corta solamente una secuencia específica de bases nitrogenadas en el ADN, siempre y cuando éstas no estén protegidas (metiladas). Por consiguiente cualquier ADN que no esté metilado puede ser reconocido y cortado en fragmentos de longitud definida; y cualquier mutación dentro de esos sitios, podría cambiar el patrón del fragmento y permitir que se detecte un RFLP al comparar dos o más genomas (Valadez y Kahl, 2000). Para detectar RFLPs hay dos técnicas: southern blots e hibridización, y PCR.

Southern blots e hibridización

Es necesario en primer lugar aislar el ADN del organismo de interés (véase el capítulo 15 de este libro, purificarlo y cortarlo con una o más endonucleasas de restricción. Los fragmentos resultantes se separan en geles de agarosa por electroforesis y se transfieren a una membrana que puede ser de nylon o de nitrocelulosa (southern blotting). La subsecuente hibridación con una sonda marcada y detección con autorradiografía, luminografía o quimioluminiscencia hará visible un fragmento de ADN específico, que puede ser comparado con el fragmento resultante de otros genomas tratados de la misma forma (Parket *et al*., 1998). La sonda marcada se obtiene de fragmentos de ADN nuclear, de organelos o bien de ADN complementario, también denominado ADN copia (ADNc) que es una molécula de una sola hilera producida en el laboratorio usando mRNA como molde y transcripción reversa (véase el capítulo 15 de este

libro). Para el caso de obtenerla a partir de ADN nuclear, se requiere construir una biblioteca genómica mediante el aislamiento, restricción y clonación del ADN en un vector apropiado (Valadez y Kahl, 2000), que se explica a continuación. Las bibliotecas de ADN son colecciones de vectores recombinantes, que cuentan con secuencias de ADN insertadas. La figura 5 muestra la construcción de una biblioteca de ADNc en el vector lambda gt10. En la figura se puede observar cómo, mediante ADN transcriptasas, polimerasas o ligasas se construye un fragmento de ADN (el de interés) y se inserta en el fago a utilizar (parte A de la figura); una vez obtenido el fago se procede a infectar una colonia de *E. coli* en una caja de Petri. Cada infección producirá una placa característica en la caja de Petri (en una caja de Petri de 150 mm pueden caber 50 mil placas). Una vez que las placas se han formado se coloca un filtro de nitrocelulosa en contacto con las colonias de la caja de Petri. El ADN transferido al filtro es desnaturalizado y fijado en el filtro mediante el uso de luz ultravioleta. De esta forma el filtro puede ser hibridizado con una sonda radiactiva de ADNc de interés y sometido a autorradiografia; una mancha en la autorradiografia permitirá identificar la placa de la caja de Petri que contiene el fago con el ADN de interés. El fago puede entonces ser extraído del agar de la caja de Petri y reinfectar a *E. coli.* La repetición de este proceso de separación de un fago recombinante, permite clonar el ADN que contiene el fago. Este proceso ha permitido desarrollar bibliotecas de ADN de diferentes tejidos.

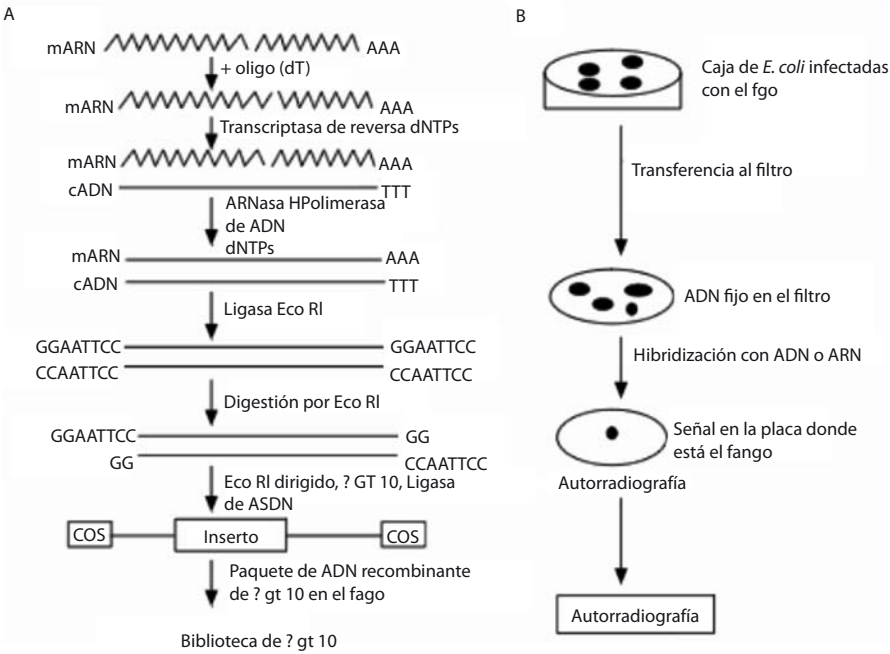
RFLP-PCR

El segundo método que se utiliza para la técnica de RFLPs es la PCR. Los RFLPs son causados por rearreglos del ADN, tales como pérdidas, inserciones o sustituciones de secuencias o nucleótidos únicos, lo que genera una ganancia o pérdida de sitios de restricción. Esto es detectado a través de diferencias en el peso molecular de los fragmentos homólogos de restricción del ADN genómico (figura 6).

Con el producto amplificado de PCR se realiza una digestión con diferentes enzimas de restricción. Los fragmentos resultantes de la digestión son separados por electroforesis en geles de agarosa teñidos con bromuro de etidio o con acrilamida teñidos con nitrato de plata (Karl *et al.,* 1992); este método es más barato y rápido que el anterior.

Los RFLPs generalmente se han utilizado para construir mapas genéticos, para la clonación de genes basados en mapas y para ayudar a resolver problemas taxonómicos o filogenéticos.

Figura 5. Construcción de una biblioteca de ADN en el vector lambda gt10 (A). En (B) se muestra en forma esquematizada el cernimiento para obtener la clona pura

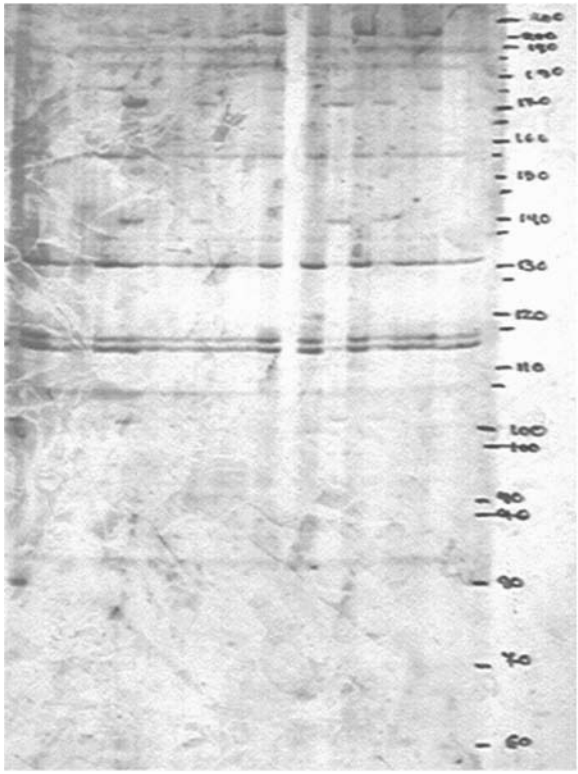


El grado de polimorfismo detectado con esta técnica difiere ampliamente entre las especies dependiendo de la sonda utilizada. Entre las principales desventajas que presenta está el requerimiento de grandes cantidades de ADN de buena calidad para la detección de loci de copias únicas, además requiere de muchas manipulaciones y solamente se detecta una fracción de la variabilidad de secuencias existentes en el genoma, es decir, su información es limitada.

AFLPs

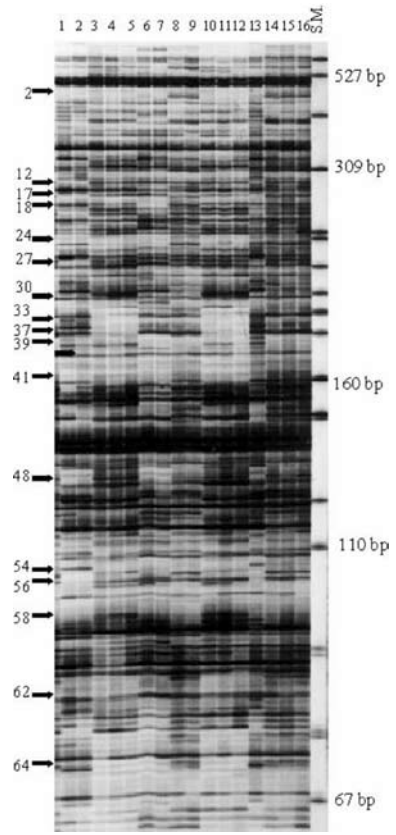
Los AFLPs (polimorfismos de longitud de los fragmentos amplificados) son una técnica que combina la digestión de dos enzimas de restricción, generalmente Mse I que reconoce y corta 4 pb y Eco RI que reconoce y corta 6 pb dentro de una secuencia, con la unión de secuencias específicas al nucleótido ligado a los extremos de los fragmentos de restricción y dos

Figura 6. Patrones de PCR-RFLPs del intrón de la mitocondria NAD1B2f y NAD1C1r digerido con la enzima Hae III en *Pinus pinceana* en gel de acrilamida al 5% y tinción con nitrato de plata (foto de Miroslava Rentería Alcántara)



amplificaciones de PCR usando primers marcados basados en las secuencias ligadas. La primera amplificación se lleva a cabo con el uso de iniciadores que contienen una base extra en el extremo 3', lo que produce un conjunto de fragmentos que además de llevar la secuencia complementaria al iniciador, complementan a la base extra adicionada. Posteriormente estos productos amplificados sirven para hacer la segunda amplificación en la que se consideran iniciadores con dos o tres bases extras. Los fragmentos resultantes son complementarios, además del iniciador, a las extensiones consideradas, razón por la cual sólo una porción del genoma fragmentado es finalmente amplificado (Vos *et al*., 1995).

Figura 7. Patrones de AFLPs en *Allium sativum* en gel de acrilamida y tinción con nitrato de plata (foto de Meryem Ipek)



Esta técnica detecta múltiples loci polimórficos y es útil para generar huellas genéticas y mapeo; también se ha utilizado para la caracterización de germoplasma, estudios filogenéticos en plantas, bacterias, hongos y en estudios de genética de poblaciones.

Entre las ventajas que se pueden encontrar en los AFLPs está que no requieren ninguna información previa de la secuencia para su análisis; se producen una gran cantidad de bandas polimórficas; la técnica es altamente reproducible y existen kits estandarizados. Sin embargo, requiere gran número de pasos para producir resultados.

Aunque actualmente el equipo es sofisticado y se necesita la radiactividad para generar los AFLPs, el método se puede automatizar fácilmente para el

alto rendimiento del procesamiento de muestras. Las marcas fluorescentes están substituyendo a la radiactividad y el número de datos producidos en un tiempo corto compensa los costos.

Las bandas observadas en los geles de AFLPs son clasificadas como presencia o ausencia de cada individuo y el análisis se desarrolla como un sistema dominante recesivo (Simpson, 1997).

Secuenciación de ADN

Los análisis más detallados de diferenciación de ADN pueden obtenerse secuenciando la región de interés para diferentes individuos. Hasta hace poco tiempo, el uso extenso de secuencias de ADN para los estudios de poblaciones no habían sido prácticos porque los fragmentos de ADN para cada individuo tenían que ser aislados para librerías de ADN subgenómico después de haber sido identificados por *southern blotting* e hibridización. Sin embargo, actualmente el uso de secuencias es muy común: se ha aplicado en análisis de genética de poblaciones y problemas taxonómicos.

Este método incluye cuatro pasos: i) identificar secuencias que tenga la variación necesaria; ii) aislar y purificar un número elevado de la secuencia (ya sea por clonación o amplificación); iii) secuenciar; iv) alinear la secuencia (con los programas MALIGN, Clustal V para Macintosh ver. 1.5, JACK ver. 4.2).

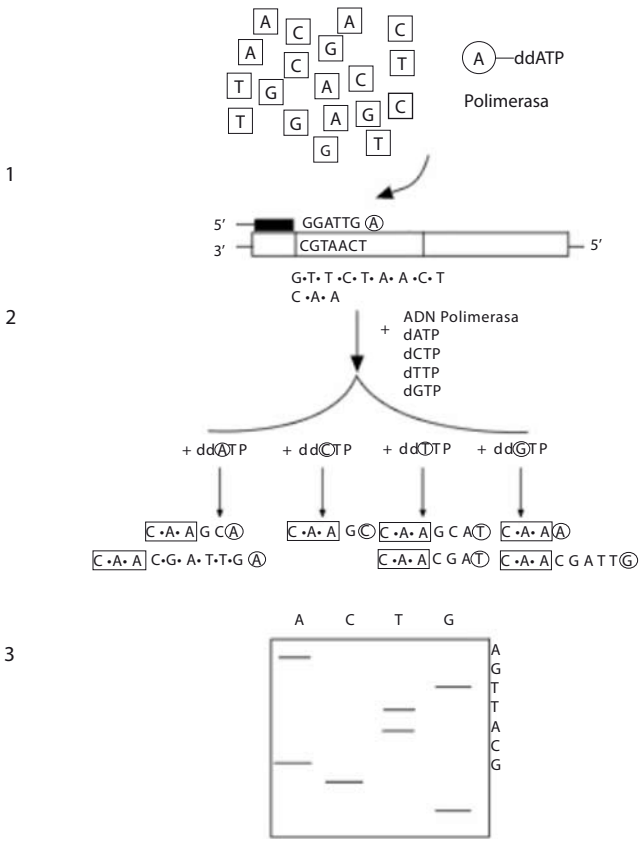
Las técnicas que existen para secuenciar son la química, la enzimática y la automática.

La química diseñada por Alan Maxan y Walter Gilbert consiste en dividir el ADN en cuatro muestras y tratar cada una con agentes químicos que rompen el ADN específicamente en una base, bajo condiciones en las que solo unos pocos nucleótidos del fragmento se afectan. Los fragmentos se marcan con radiactividad y se auto radiografían.

En el método enzimático de Fred Sanger el ADN a ser secuenciado se utiliza como patrón de referencia para producir una síntesis *in vitro*, mediante una polimerasa y una réplica de ADN que inicia la síntesis del ADN siempre en el mismo sitio. El punto crítico es el uso de trifosfato de dideoxirribonucleósido: la deoxirribosa no cuenta con el grupo 3-OH, de tal forma que cuando el nucleótido es incorporado a la cadena de ADN, se bloquea la adición del siguiente nucleótido, por lo que cada molécula de ADN sintetizada va a terminar aleatoriamente en dicho nucleótido. Esta síntesis se lleva a cabo en cuatro muestras; en cada muestra se incorpora un dideoxinucleósido con una base diferente (ddATP, ddCTP, ddTTP, ddGTP) y los otros tres deoxinucleótidos

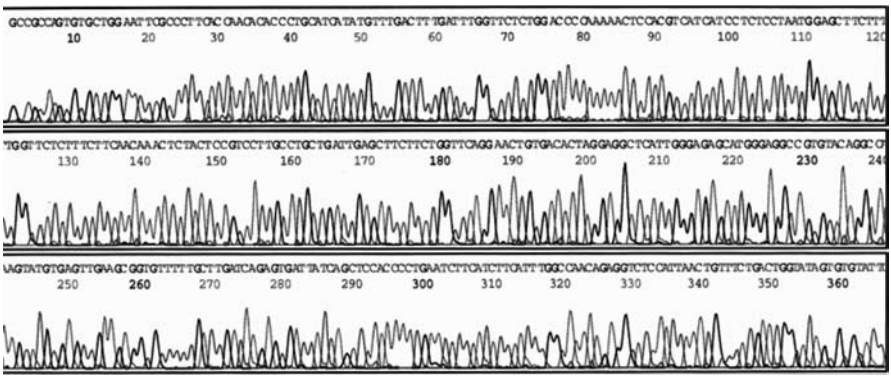
normales. Los fragmentos sintetizados durante el proceso van a terminar en sitios diferentes de la cadena del ADN, generando una escalera de moléculas de ADN que pueden ser detectadas por auto radiografía. Las cuatro muestras se corren en electroforesis en líneas paralelas para determinar el orden de los nucleótidos, identificando la secuencia de ADN que será complementaria al ADN que sirvió de modelo para la síntesis (figura 8).

Figura 8. Método se secuenciación enzimática de ADN. 1) Síntesis en presencia del ADN de interés, polimerasa, un fragmento de ADN como iniciador, nucleótidos y dideoxinucleótidos atípicos; 2) Síntesis por separado para cada nucleótido, generando fragmentos de ADN con el dideoxinucleótido incorporado; 3) separación de los fragmentos en un gel para identificar la secuencia



Recientemente se ha introducido la secuenciación automática que utiliza la reacción de Fred Sanger pero con fluorescencia en vez de radiactividad, que se detecta por medio de un láser. Cada emisión fluorescente es transmitida como una señal directamente a la computadora donde un programa la interpreta y la codifica como un nucleótido particular (Swofford *et el*., 1996; Griffiths *et al*., 1996; Ferl *et al*., 1991). La captura de secuencias se hace más rápida debido a que incorporan programas que leen los nucleótidos directamente desde el gel (figura 9). Entre las principales ventajas que tiene la secuenciación de ADN es su alta reproducibilidad y que es codominante; su mayor limitante es su alto costo pero es la mejor alternativa entre los métodos con isótopos radiactivos.

Figura 9. Ejemplo de patrones de secuenciación automática



Microarrays (microarreglos)

El microarray es un arreglo de cientos de millares de secuencias de ADN inmovilizadas y bien ordenadas en forma de matriz adheridas a una superficie sólida, generalmente de cristal. Cada secuencia corresponde a un gen diferente. Esta técnica es utilizada para detectar niveles de expresión de los genes colocados en el microarray. Esto se logra extrayendo el mARN de la célula de interés o de algún tejido particular, sometiéndolo a transcripción inversa para obtener cADN (véase el capítulo 16 de este libro)y combinando éste con un marcador fluorescente (usualmente los tintes de cianina Cy3 y Cy5) para generar una sonda, que es hibridizada sobre la matriz del microarray, de manera que para cada gen se detecta el grado de fluorescencia, lo cual da el nivel de expresión del gene.

Los microarrreglos pueden ser divididos en dos tipos principales que difieren en su construcción: spotted microarrays (microarreglos punteados) y arreglos de oligonucleótidos de alta densidad. Los microarreglos punteados usan generalmente muestras de ADN de 500 a 5000 bases (Ekins *et al*., 1999) producidos por PCR a partir de bibliotecas o de bases de datos como GenBank, UniGene, dbEST, etc. Las secuencias son depositadas sobre la superficie sólida por un robot en lugares definidos. Hay dos métodos de depositar los puntos:

Distribuidor activo: basado en la tecnología de las impresoras de inyección de tinta.

Distribuidor pasivo: aplica la solución de ADN con un alfiler que toca la superficie sólida. Se puede obtener una mayor densidad de manchas usando este método.

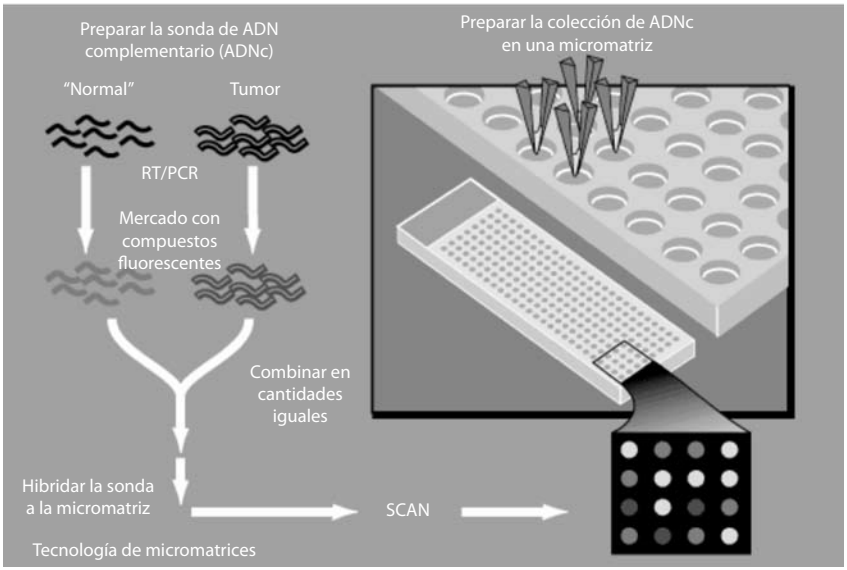
La superficie sólida generalmente es un portaobjetos especialmente cubierto pero también se pueden usar membranas de nylon y portaobjetos cubiertos de oro. El tamaño de los puntos de la matriz típicamente varía de 80 a 150 μm de diámetro y en un solo portaobjetos cabe un máximo de 80,000 puntos. En los microarrays punteados es común comparar las expresiones genéticas de dos muestras biológicas (tejido con tumor vs. tejido sin tumor, por ejemplo) en un mismo portaobjetos, así que el mRNA es preparado para que los dos niveles de expresión puedan ser medidos (Schena *et al*., 1995; figura 9).

Por su parte, los microarreglos de oligonucleótidos de alta densidad son construidos comercialmente y tienen una precisión y densidad muy alta al usar oligonucleótidos cortos de 20 a 25 bases.

Dos de las compañías que construyen estas clases de micrroarreglos son Affymetrix y Agilent Technologies. Los microarreglos de Affymetrix son los más populares y son conocidos como GeneChipsTM, producidos sintetizando decenas de miles de oligonucleótidos cortos *in situ* usando técnicas de fotolitografía. Affymetrix produce diferentes microarrays, cada uno con diferente composición de genes (Lockart *et al.*, 1996).

Los microarreglos necesitan ser leídos por un instrumento apropiado para convertir la distribución de florescencia o radioactividad en una imagen de computadora. En el caso de las sondas fluorescentes, se pueden usar escáners de láser o cámaras CCD. En la imagen de computadora cada intensidad del punto corresponde a un nivel de expresión de la sonda. Puesto que un punto contiene múltiples pixeles, es normal promediar todos los pixeles de un mismo punto junto con los pixeles del borde exterior. Sin embargo, la cuantificación

Figura 10. Técnica de microarrays para detectar niveles de expresión de genes (tomada de la página web del Instituto de Neurobiología, UNAM)



del grado de fluorescencia está sujeta a las fluctuaciones de intensidad dentro de cada punto, aunque éstas pueden ser superadas con las técnicas desarrolladas por Brown *et al*. (1999).

Los microarreglos de Affymetrix presentan problemas similares, pero también se han desarrollado algoritmos para combatirlos.

Esta tecnología permite a los investigadores estudiar la relación de genes individuales de enfermedades (por ejemplo el cáncer), y después aplicar el conocimiento al tratamiento de pacientes.

También se ha utilizado en el estudio de tejidos finos, en células normales durante el desarrollo y estudios de animales transgenicos.

La tecnología de los microarreglos también tiene uso potencial en el desarrollo de nuevas drogas así como en la investigación de farmacogenómicos (cuyo fin es encontrar la correlación entre las respuestas terapéuticas a las drogas y los perfiles genéticos de los pacientes) y de toxicogenómicos (cuya meta es encontrar correlaciones entre los toxicantes y los cambios en los perfiles genéticos de los objetos expuestos a los toxicantes; Leming Shi, 1998, 2002).

Los dos grandes problemas que se presentan con los datos de microarreglos son por un lado los niveles de expresión, que varían de experimento a experimento, y por las diversas fuentes de errores aleatorios y sistemáticos durante el análisis. Por otro lado, hay un pequeño número de muestras comparado con el gran número de variables, lo que causa que las técnicas estadísticas tradicionales fracasen. Otra gran limitante de los microarrays proviene, irónicamente, de su productividad notable, y actualmente los científicos apenas pueden hacer frente al volumen enorme de información producida por los microarreglos, además de tener un alto costo.

Bibliografía

Aagaard J.E., V.K. Krutovskii y H.S. Strauss. 1998. RAPDs and allozymes exhibit similar levels of diversity and differentiation among populations and races of Douglas-fir. *Heredity* 81:69-78.

Armour A. L., R. Neuman, S. Gobert y A.J. Jeffreys. 1994. Isolation of human simple repeat loci by hybridization selection. *Human Molecular Genetics* 3(4):599-605.

Ashley C.T. y S.T. Warren. 1995. Trinucleotide repeat expansion and human disease. *Annual Review Genetics* 29:703-728.

Ayala F.J. 1984. *Genética Moderna*. Fondo Educativo Interoamericano, Barcelona.

Avise, C.J. 1994. *Molecular markers, natural history and evolution*. Chapman & Hall, New York.

Backeljau T., L. De Brun, H. De Wolf, K. Jordaens, S. Dongen, R. Verhagen y B. Winnepenninckx. 1995. Random amplified polymorphic DNA (RAPD) and parsimony methods. *Cladistics* 11:119-130.

Berovides V. y M.A. Alfonso. 1995. *Biología evolutiva*. Pueblo y Educación, La Habana.

Brettschneider R. 1998. *Molecular tools for screening biodiversity in plants and animals*. Chapman & Hall, Londres.

Brown A.H.D. 1979. Enzyme polymorphisms in plant populations. *Theoretical Population Biology* 15:1-42.

Brown W.M., M. George Jr. y A. C. Wilson. 1979. Rapid evolution of mitochondrial DNA. *Proc. Natl. Acad. Sci. USA* 76:1967-1971.

Brown, P.O. y D. Botstein. 1999. Exploring the new world of the genome with DNA microarrays. *Nature Genetics* 21: 33–37.

Coltman D.W., D.W. Bowen y J.M. Weith. 1996. PCR primers for harbour seal (*Phoca vitulina concolour*) microsatellites amplify polymorphic loci other pinniped species. *Molecular Ecology* 5:161-163.

Culley M.T. y A.D. Wolfe. 2001. Population genetic structure of the cleistogamous plant species *Viola pubescens* Aiton (Violaceae), as indicated by allozyme and ISSR molecular markers. *Heredity* 86:545-556.

Dervey M. E., J.C. Bell, D.N. Smith, D.B. Neale y G.F. Moran. 1996. A genetic linkage map for *Pinus radiata* based on RFLP, RAPD and microsatellite markers. *Theor. Appl. Genet*. 92:673-679.

Dirienzo A. y C. Wilson. 1991. Branching pattern in the evolutionary tree for human mitochondrial DNA. *Proc. Natl. Acad. Sci. USA* 88:1597-1601.

Echt C. S., L. De Verno, M. Arzide y G.G.Vendramin. 1998. Chloroplast microsatellites
reveal population genetic diversity in red pine, *Pinus resinosa. Molecular Ecology* 7:307-309.

Eduars A., A. Civetello, H.A. Hammond y T. Caskey. 1991. DNA typing and genetic mapping with trimeric and tetrameric tandem repeats. *American Journal Human Genetics* 49:746-756.

Ekins R. y F.W. Chu. 1999. Microarrays: their origins and applications. *Trends in Biotechnology* 17:217-218.

Golstein D.B. y D.D. Pollok. 1994. Least-squares estimation of molecular distance – noise abatement in phylogenetic reconstruction. *Theoretical Applied Genetics* 12:432-440.

Golstein D.B., L.A. Ruíz, L.L. Cavalli-Sforza y M.W. Feltman. 1995a. An evaluation of genetic distances for use with microsatellite loci. *Genetics* 139:463-471.

Golstein D.B., L.A. Ruíz, L.L. Cavalli-Sforza y M.W. Felman. 1995b. Genetics absolute dating based on microsatellites and the origin of modern human. *Proc. Natl. Acad. Sci. USA* 92:6723-6727.

Golstein D.B., L.A. Zhivotovsky, K. Nayar, L.A. Ruíz, L.L. Cavalli-Sforza y M.W. Feltman. 1996. Stadistical properties of the variation at linked microsatellite loci: implications for the history of human Y-chromosome. *Molecular Biology and Evolution* 13:1213-1218.

Griffiths A.J.F., J.H. Miller, D.T. Suzuki, R.C. Lewontin y W.M. Gelbart 1993. *Genetic Analysis*. Freeman. New York.

Gupta M., Y.S. Chyi, J. Romero-Severson y J.L. Own. 1994. Amplification of DNA markers from evolutionary diverse genomes using single primers of simple-sequence repeats. *Theoretical. Applied Genetics* 89:998-1006.

Gwynne P. y G. Page. Microarray analysis: the next revolution in molecular biology. *Science*. 1999. Special advertising supplement.

Harris H. 1966. Enzyme polymorphisms in man. *Proceedings of the Royal Society* Series B 164:298-31.

Hillis D.M. 1987. Molecular versus morphological approaches to systematics. *Annual Review of Ecology and Systematics* 18:23-42

Hillis D.M. 1994. Homology in molecular biology. *In*: B. K. Hall (ed.). Homology: the hierarchical basic of comparative biology. Academic Press, San Diego. pp. 339-368.

Hillis D.M., B.K. Mable, A. Larson, S.K. Davis y E.A. Zimmer. 1996. Nucleic Acids IV: sequencing and cloning. *In*: Hills D.M., C. Moritz, B.K. Mable. *Molecular Systematics*. Sinauer Associates Inc. Massachusetts. pp. 321-378.

Hillis D.M., M.T. Dixon. 1991. Ribosomal DNA : molecular evolution and phylogenetic
inference. *Quarterly Review of Biology* 66(4):411-453.

Hillis D.M., J.J. Wiens. 2000. Molecules versus morphology in systematics. *In*: J.J. Wiens (ed.). Phylogenetic analysis of morphological data. Smithsonian Institution Press. Washington. pp. 1-19.

Hubby J.L. y R.C. Lewontin. 1966. A molecular approach to the study of genio heterozygosity in natural populations. The number of alleles at different loci in *Drosophila pseudoobscura*. *Genetics* 54:577-59.

Johnson F.M., C.G. Kanapi, R.H. Richardson, M.R. Wheeler y W.S.Stone. 1966. Analysis of polymorphisms among isozyme loci in dark and light *Drosophila ananassae* strains from America and Western Samoa. *Proc. Natl. Acad. Sci. USA* 56:119-125.

Karl S.A. y C. Avise. 1992. Balancing selection at allozyme loci in oysters, implications from nuclear RFLP's. *Science*. 256:100-1001.

Karl S.A., B.W. Bowen y J.C. Avise. 1992. Global population genetic structure and male-mediated gene flow in the green turtle (Chelonia mydas): RFLP analysis of anonymous nuclear loci. *Genetics* 131:163-173.

Kochert G. 1991. Restriction fragment lenght polymorphism in plants and its implications. *In*: Biswas B. B., J.R. Harris. Subcellular biochemistry, vol. 17: plant genetic engineering. Plenum Press, New York. pp. 167-190

Litt M. y J.A. Luty. 1989. A hypervariable microsatellite revealed by in vitro amplification of a dinucleotide repeat within the cardiac muscle gene. *American Journal of Human Genetics* 44:397-401.

Lynch M. y. B.G. Milligan. 1994. Analysis of population genetic structure with RAPD markers. *Molecular Ecology* 3:91-99.

Lockhart D.J. 1996. Expression monitoring by hybridization to high-density oligonucleotide arrays. *Nature Biotechnol*. 14:1675–1680.

Markert C.L., F. Moller 1959. Multiple forms of enzymes: tissue, ontogenic and species-specific patterns. *Proc. Natl. Acad. Sci. USA* 45:753-763.

McCracken G. F., G.M. Burghard y S.E. Houst.1999. Microsatellites marker and multiple paternity in the garter snake *Thamnophis sirtalis. Molecular Ecology* 8:1475-1479.

Ortíz A., C. Ramis, P. Parra, A. Díaz y L. López. 2002. Patrones isoenzimáticos de variedades de arroz y arroces rojos en Venezuela. *Rev. Fac. Agro.* (Maracay) 28:117-130.

Otero A.A., M. De la Cruz y K. Oyama. 1997. El uso de los RAPDs como marcadores moleculares en plantas. *Bol. Soc. Bot. Mex.* 60:85-117.

Otha T. y M. Kimura. 1973. The model of mutation appropriate to stimate the number of electrophoretically detectable alleles in a genetic population. *Genetic Research* 22: 201-204.

Parker P.G., A.A. Snow, M.D. Schug, G.C. Booton y P.A. Fuerst. 1998. What molecules con tell us about population: choosing and using a molecular marker. *Ecology* 79:361-382.

Pérez-Nasser N. y D. Piñero.1997. Isoenzimas. *Bol. Soc. Bot. Méx*. 60:77-84.

Powell W., M. Morgante y C. Andre. 1995a. Hypervariable microsatellites provide a general source of polymorphic DNA markers for the chloroplast genome. *Current Biology* 5:1023-1029.

Prakash S., R.C. Lewontin y J.L. Hubby. 1969. A molecular approach to the study or genic heterozygosity in natural populations. IV. Patterns of genic variation in central marginal and isolated population of *Drosophila pseudooscura*. *Genetics* 61:841-858.

Provan J., N. Soranzo, N.J. Wilson, D.B. Gostein y W. Powell. 1999. A low mutation rate for chloroplast microsatellites. *Genetics* 153:943-947.

Queller D.C., J.E. Strasmann, C.R. Hughes. 1993. Microsatellites and kinship. *Trends Ecol. Evol.* 8:285-289

Reiter R.S., J.A. Rafalski, S.V. Tingey y P.A. Scolnick 1992. Global and local genome mapping in *Arabidopsis thaliana* by using recombinant inbred lines and random amplified polymorphic DNA. *Proc. Natl. Acad. Sci. USA* 89:1477-1481.

Rieseberg L.H. y D.E. Soltis. 1991. Phylogenetic consequences of cytoplasmic gene flow in plants. *Evol. Trends Plants* 5: 65-84.

Russell J.R., J.D. Fuller, M. Macaulay, BG. Hatz, A. Jahoor, W. Powell y R. Waugh. 1997. Direct comparison of leves of genetic variation among barley accessions detected by RFLPs, AFLPs, SSRs and RAPDs. *Theor. Appl. Genet*. 95:714-722.

Sanderson M.J. y M. J. Donoghue.1989. Patterns in levels of homoplasy. Evolution 44:1671-1684.

Schena M., D. Shalon, R.W. Davis y P.O. Brown. 1995, Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science* 5235:467-70.

Shriver M.D. 1995. A novel measure of genetic distance for hightly polymorphic tandem repeat loci. *Mol. Biol. Evol.* 12: 914-920.

Simpson J. 1997. Amplified fragment length polymorphisms. *Bol. Soc. Bot. Méx.* 60:73-76.

Soranzo N., J. Porvan. y W. Powell. 1999. An example of microsatellite length variation in the mitochondrial genome of conifers. *Genome* 42:158-161

Slatkin M. 1995a. A measure of population subdivision based on microsatellite allele frequencies. *Genetics* 139:457-462.

Stansfield W.D. 1992. *Genética*. McGraw- Hill, México.

Swofford D.L., G.J. Olsen, P.J. Waddell y D.M. Hillis. 1996. Phylogenetic inference. In: Hillis D.M., C. Moritz y B.K. Mabbe. *Molecular Systematics*. Sinauer Associates Inc. Massachusetts. pp. 407-510

Valadez E. y G. Kahl. 2000. *Huellas de ADN en genomas de plantas*. Universidad Autónoma Chapingo. México.

Vendramin G.G., L. Lelli, P. Rossi y M. Morgante. 1996. A set of primers for the amplification of 20 chloroplast microsatellites in Pinaceae. *Molecular Ecology* 5:595-598.

Vendramin G.G., M. Anzide, Madaghiele y G. Bucci. 1998. Distribution of genetic diversity in *Pinus pinaster* Ait. as revelad by chloroplast microsatellites. *Theor. Appl. Gen.* 97: 456-463.

Vos P., R. Hogers, M. Bleeker, M. Reijans, T. Van de Lee, M. Hornes, A. Frijters, J. Pot, J. Peleman, M. Kuiper y M. Zabeau. 1995. AFLP: a new technique for DNA fingerprinting. *Nucleic Acids Research* 23:4407-4414.

Weber L. y C. Wong. 1993. Mutation of human short tandem repeats. *Human Molecular Genetics* 2:1123-1128.

Whitkus R., J. Doebley y J.F. Wendel. 1994. Nuclear DNA markers in systematics and evolution. *In*: Phillips R.L. y J.K. Vasil (ed.). *DNA-based markers in plants*. Klumer Academic Publishers, Netherlands. pp. 116-141.

Wiessenbach, J. y G. Dib. 1992. A second-generation link-age map of the human genome. *Nature* 359: 794-801.

Williams J.G.K., A.R. Kubelik, K.J. Livak, J.A. Rafalski y S.V. Tingey. 1990. DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. *Nucl. Acids Res*. 18:6531-6535.

Zhivotovsky L.A. 1999. Estimating population structure in diploids with multilocus dominant DNA markers. *Molecular Ecology* 8:907-913.

Capítulo 19

Inter Simple Sequence Repeats (ISSRs)

Andrea González y Xitlali Aguirre

Los ISSRs son un tipo de marcador genético que nos permite obtener los niveles de variación en las regiones microsatélite que se encuentran dispersas en varios genomas, particularmente el nuclear. Estas regiones consisten en repeticiones en tandem de motivos simples como $(CT)_n$ ó $(CA)_n$, ubicadas entre secuencias no repetitivas del genoma nuclear eucarionte. Los motivos repetidos, llamados también SSRs (simple sequence repeats) pueden ser penta-, tetra-, tri- y dinucleótidos. La longitud de las secuencias de microsatélites tiende a ser altamente variable entre individuos debido a las altas tasas de mutación que experimentan, ya que cuando el DNA se replica durante la meiosis, la DNA polimerasa puede "tartamudear" hacia adelante o hacia atrás en las unidades repetidas, eliminando o agregando unidades a la cadena. Las cadenas resultantes pueden entonces presentar menos o más unidades de repetición (o pares de bases) que las cadenas parentales (Zietkiewicz *et al*., 1994; Wolfe, 2000).

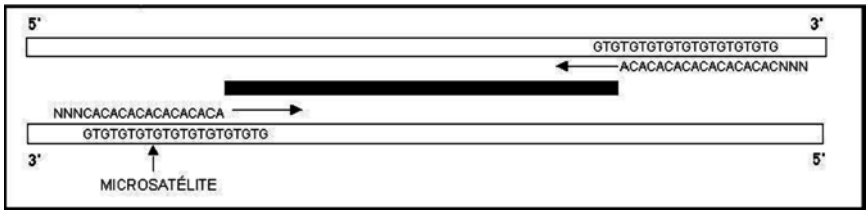
Los ISSRs son marcadores semiarbitrarios amplificados por la reacción en cadena de la polimerasa (PCR) a partir de la presencia de un oligonucleótido o primer complementario a un microsatélite, diseñado para unirse a los motivos repetidos de di y trinucleótidos (evitando los mononucleótidos presentes en el cloroplasto). Los primers de ISSRs consisten en un motivo repetido de di- o trinucleótidos complementario a la secuencia del microsatélite. En ocasiones es posible agregar a esta secuencia un par de nucleótidos extras arbitrarios en

el extremo 3' o en el 5', que jugarán el papel de "anclas" asegurando así que la amplificación inicie siempre en el extremo 5' o en el 3' del microsatélite, respectivamente (Zietkiewicz et al., 1994; Bornet y Branchard, 2001; Pradeep, 2002).

Cuando dos secuencias repetidas se presentan dentro de una distancia amplificable y con una orientación invertida, el primer complementario a ellas (uno por reacción de PCR; véase el capítulo 17 de este libro) puede inducir la amplificación del segmento de ADN intermedio.

La molécula generada, con un tamaño particular (peso molecular), se considera un "locus", que representa el segmento de ADN entre los microsatélites. Y se ha visto que los ISSRs frecuentemente amplifican de 25 a 50 bandas en una sola reacción. Este patrón característico de productos de PCR se considera la "huella digital genética" de cada uno de los individuos analizados. El polimorfismo entre individuos de la misma población puede detectarse, ya que el análisis es sensible a la presencia/ausencia del elemento genómico reconocido por el primer y a la longitud de la secuencia intermedia amplificada (Zietkiewicz et al., 1994).

Figura 1. Amplificación con un primer $(CA)_n$ anclado en el extremo 5' con tres nucleótidos extras. Se amplifica el segmento intermedio entre dos secuencias de microsatélite en orientación invertida



Las bandas de ISSRs son consideradas marcadores dominantes. La presencia de la banda representa el genotipo dominante (homócigo o heterócigo), mientras que su ausencia representa el genotipo homócigo recesivo. Se asume que existen dos alelos por locus.

La ausencia de una banda puede deberse a varios factores:

1) La no existencia de un sitio de unión completo al primer debido a una mutación,
2) Rearreglos estructurales en el cromosoma durante la meiosis,
3) Inserciones o deleciones suficientemente grandes como para aumentar o

disminuir el tamaño de la banda, de manera que se identifica como un locus diferente.

Ventajas y desventajas

Las ventajas que ofrece esta técnica se centran principalmente en la alta variación que detecta, así como en su reproducibilidad debida principalmente a las altas temperaturas de alineación utilizadas en la PCR. Asimismo, no son necesarias altas concentraciones de ADN. Por otro lado, para diseñar los primers no es necesario conocer la secuencia del genoma del organismo en estudio. Pueden visualizarse tanto en geles de agarosa como de acrilamida. Finalmente, son sencillos de montar, rápidos, eficientes y poco costosos.

En cuanto a las desventajas, la homología de las bandas es incierta. Y dado que son marcadores dominantes, no permiten el cálculo de ciertos parámetros que exigen distinguir a los heterócigos de los homócigos dominantes (como F_{IS} y F_{IT}; Wright, 1965). Asimismo, para estimar la heterocigosis poblacional es necesario asumir *a priori* que la población se encuentra en equilibrio de Hardy-Weinberg, por lo que las estimaciones de heterocigosis y estructuración genética pueden sesgarse un poco, aunque existen correcciones estadísticas (Lynch y Milligan, 1994).

Aplicaciones

Los ISSRs pueden ser utilizados como marcadores en la resolución de múltiples problemas biológicos. Los polimorfismos que presentan, además de su heredabilidad, permiten aplicarlos en la identificación de individuos, distinción de variedades intraespecíficas (particularmente en especies con importancia económica), identificación de paternidad y maternidad, mapeo genético, evaluación de diversidad y subdivisión genética en poblaciones, reconstrucciones filogenéticas, introgresión e hibridización, y distinción de individuos con origen clonal y sexual (Zietkiewicz *et al*., 1994; Wolfe, 2000; Pradeep *et al*., 2002).

Ejemplo

Al empezar a trabajar con ISSRs es recomendable probar diferentes primers en una condición estándar y después optimizar aquellos que dieron resultados positivos. Esta optimización incluye modificar las condiciones iniciales de

la PCR, particularmente la temperatura y tiempo de alineación, así como la concentración de $MgCl_2$.

Por ejemplo, en algunos estudios de genética de poblaciones con especies del género *Agave* (Aguirre, 2004; González, 2004) se amplificaron ISSRs con los siguientes primers:

841	GAG AGA GAG AGA GAG AYC	Y=C o T
842	GAG AGA GAG AGA GAG AYG	R=A o G
846	CAC ACA CAC ACA CAC ART	
857	ACA CAC ACA CAC ACA CYT	

Para una reacción de 25 µl se agregaron 2.5 µl de amortiguador para PCR 10X, 0.5 µl de dNTP mix (mezcla de los 4 dideoxinucleótidos a 10 mM), 0.2 µl de Taq polimerasa a 5 U/µl; 2 µl de DNA a 15 ng/µl; el volumen de $MgCl_2$ (concentración inicial de 30 mM) depende de la concentración óptima para cada primer y para completar los 25 µl de la reacción, se agregó el volumen necesario de agua ultrapura correspondiente (tabla 1)

Tabla 1. Condiciones óptimas de amplificación para algunos primers.

Primer	Temperatura de alineación	Tiempo de Alineación	[$MgCl_2$] (mM)	No. de ciclos
841	60° C	45"	2.1	30
842	60° C	35"	2.3	35
846	52° C	45"	0.96	30
857	52° C	45"	1.5	30

Una vez que se visualizaron las bandas en geles de agarosa al 2% teñidos con bromuro de etidio, se hizo un listado de todas las bandas o loci ampificados en la población. Con estos datos se elaboró una matriz de presencia/ausencia de las bandas por individuo. Esto puede hacerse a mano, o bien por medio de programas como el 1D 3.5 de Kodak. Una vez obtenida la matriz, los datos permitieron obtener frecuencias alélicas, y parámetros clásicos de la genética de poblaciones como heterocigosis poblacional, por especie, estimaciones de la estructuración, distancias genéticas, árboles de similitud y análisis jerárquicos de la varianza molecular. Estos cálculos pueden realizarse con diversos programas como el TFPGA (Miller, 1997) o el Arlequin (Schneider *et al.*, 2000).

Bibliografía

Aguirre X. 2004. Genética de poblaciones de *Agave cupreata* y *Agave potatorum*: aportaciones para el manejo y la conservación de dos especies mezcaleras. Tesis de licenciatura, Facultad de Ciencias, UNAM. 73 pp.

Bornet B. y Branchard M. 2001. Nonanchored inter simple sequence repeat (ISSR) markers: reproducible and specific tools for genome fingerprinting. *Plant Molecular Biology Reporter* 19:209-215

González A. 2004. Biología reproductiva y genética de poblaciones del *Agave garciae-mendozae*. Tesis de licenciatura, Facultad de Ciencias, UNAM. 88 pp.

Lynch M. y Milligan B.G., 1994. Analysis of population genetic structure with RAPDs markers. *Molecular Ecology* 3:91-99

Miller M.P., 1997. Tools For Population Genetic Analyses (TFPGA) 1.3: A Windows program for the analysis of allozyme and molecular population genetic data. Computer software distributed by author.

Pradeep R., Sarla N. y Siddiq E.A., 2002. Inter simple sequence repeats (ISSR) polymorphism and its aplication in plant breeding. *Euphytica* 128(1):9-17

Schneider S., Roessli D., y Excoffier L., 2000. Arlequin ver. 2.000. A software for population genetics data analysis. Genetics and Biometry Laboratory, University of Geneva, Switzerland.

Wright S., 1965. The interpretation of population structure by F-statistics with special regard to systems of mating. *Evolution* 19:395-420

Wolfe A., 2000. ISSR Resource Website. http://www.biosci.ohio-state.edu/~awolfe/ISSR/ISSR.html

Zietkiewicz E., Rafalski A. y Labuda D., 1994. Genome fingerprinting by simple sequence repeats (SSR)-anchored polymerase chain reaction amplification. *Genomics* 20:176-183

Índice analítico

12s, genes mitocondriales, 450
16S, 401, 404, 408; genes mitocondriales, 450; rARN, 338, 352, 360, 403-406, 456
5.8 rADN, 544
18 rADN, 544
28 rADN, 544
θ, 305
χ-cuadrada, 30, 475, 476

A

A. flaviceps, 483
abundancia, distribución de, 411; modelos de, 411
ACE, estimador, 413
ácido, clorhídrico, 500; nucléico, 89, 130, 143, 398-399, 500, 501, 502; graso, 398; húmico, 501
acrilamida, 532, 533; geles de, 549, 550
adaptación molecular, 14, 395
adecuación, 230; estimación de la, 173; poblacional, 163; tasa de, 199
adenovirus, células de, 499
aditividad, condición de, 97
ADN, 16, 18, 22, 32, 49, 74, 113, 282, 284, 285, 290, 332, 355, 359, 399, 403, 405, 406, 414, 442, 473, 500, 501, 503-508, 510, 511, 513, 517, 518, 520, 526, 531, 533, 535, 543, 547, 548, 557, 569; bibliotecas de, 553; cadenas de, 165; calidad del, 528; citoplásmico, 446; como molécula informativa, 467; de cloroplasto, 435, 543; en animales, 502, 511; en bacterias gram negativas, 502; en plantas, 502, 506; *fingerprinting*, 164; fósil, 95; genómico, 518; heteroduplex, 359; hibridación, 330; microarreglos de, 165; mitocondrial (mtADN), 444, 447 469, 470, 474, 480, 481, 482, 453, 543; mitocondrial

ADN *(continúa)*
y tiempo de coalescencia del, 336; mitocondrial, variación del, 482; nuclear, 337, 548, 552 perfil de, 164; precipitación con etanol absoluto, 502; ribosomal (RADN), 544; secuencias de, 138; típicamente bacteriano circular, 543; y polen de frutos silvestres, 189; y proteínas, 12. *Véase* también microarreglos.
ADNasa, 513, 515, 523, 532
ADNc, 552
Aesculus, 434
Afipia, 353
AFLPs, 194, 198, 520, 548, 554
agamospermia, 222
agarosa, 532, 545, 552, 553
agaváceas, 506
Agave macroacantha, roseta del, 223; y pseudoviviparidad, 222, 507, 570
Agelaius phoeniceus, 482
Agregación Cladística de Haplotipos (CHA), 334
Agrobacterium, 353
agua, diversidad genómica en, 394; reacción de PCR, 523
aislamiento, geográfico, 428, 429; por distancia, 58, 479; reproductivo, 362, 363, 546; sexual, 367
ajolote (*Ambystoma tigrinum*), 477
Alaska, 457
alcohol etílico, 511
alelos, deletéreos, 15, 289; distribución geográfica de los, 480; infinitos, 74
algoritmo, genético, 109; MCMC, 146; alineación, 104, 124, 132; de dos o más secuencias, 108; métodos de comparación, 125; ambigua de sitios, 123; de secuencias, 145; local, 107; progresivo, 108
alineamiento, temperatura de, 527
almidón, 545
aloenzimas (electroforesis de proteínas), 430, 450, 469, 479
alopatría (aislamiento geográfico), 428
ambientes extremos, 415
aminoácidos, 18, 28, 29, 39, 41, 94, 119, 120, 121,130, 294, 499; i, 118; informativo, 132; j, 118; y programas de cómputo, 123
Ammodramus savannarum (gorrión), 445
análisis cladístico anidado. *Véase* GeoDis
análisis, biogeográfico, 330; de Agregación Poblacional (PAA), 334; de clados anidados (NCA), 477; de paternidad, 50, 548; filogenético, 333, 334, 339, 467; filogeográficos, 330, 446, 470; programas de cómputo y, 122
ancestro, 95; común, 104, 331
Andes de Colombia, distribución del carpintero arlequín, 486
androdioica, 195
anfibios, diferenciación filogeográfica de los, 452; estructura filogeográfica de los, 460; patrones filogeográficos de los, 461
angiospermas, 185, 188, 220, 434
animales e hibridación, 432
animales; formación de nuevas especies, 333; metazoarios, tamaño del genoma mitocondrial en, 543
antibióticos, resistencia a los, 360
apareamiento, sistema de, 70, 299
apomorfía, 95
aproximaciones bayesianas, 141
árbol, 138; coalescente, 446; estrella, 129;

aditivo, 91; de consenso, 97, 98; de genes, 101; de haplotipos, 448; enraizado, 92, 93, 135; aleatorizados, 147; de especies, 99; de genes, 99; filogenético, 89, 90, 94, 98, 127, 132, 148, 472; genealógico, 185; genético, 469; no enraizados, 92, 93; óptimo, 143; parsimonioso, 131, 132; ultramétricos, 92; genealógico, 102; más parsimonioso, 132, 472; verdadero, y método de distancia, 128
Archaea, 401, 415; aplicación del concepto de especie a, 338; y teoría evolutiva, 335
ARN, 134, 499, 500, 512-515; de transferencia, 499; mensajero, 499; moléculas de, 118; ribosomal, 499
ARNasas, 512, 513, 523, 532
arqueas y homoplasia, 332
Artemisia, 434
artritis de Lyme, 289
Asia, 457
asignación fraccionada, 170; y categórica, 170
Aspergillus flavus, estructura panmítica, 29; y aflatoxinas, 292; y estudios futuros de recombinación, 292
ATP, genes mitocondriales, 450
atracción de ramas largas, 144
Aulacorhychus prasinus (tucaneta esmeralda), 488
Auriparus flaviceps, secuencias de ADN mitocondrial, 482, 483
autócigo, 65
autofecundación, 185, 187, 190, 193, 194, 197
autofertilización, 184, 190
autogamia, 173
autopomorfía, 94, 95
autosomales, 67
aventón, 288
aves, 450, 452, 467; análisis de paternidad en, 164; en México, patrones filogeográficos, 461, 482; estructura filogeográfica, 460; paserinas, 482; terrestres, estudios filogeográficos comparativos de, 482
avestruces, análisis de parentesco, 172
Avise, John, 441, 442, 444, 449
azar, 448
Azorhizobium, 352

B

Bacillus, anthracis, 368; *cereus*, 368; *subtilis*, 340, transformación del, 285; *thuringiensis*, 368
bacterias, 365, 398, 401, 504; aislamiento sexual en, 359; aplicación del concepto de especie, 338; bacterias, arqueas y eucariontes, 150; estudios filogenéticos de, 556; dinámica evolutiva de las, 310; en la boca humana, 409; gram negativas, 503; intercambio genético, 358, 367; mecanismos de reproducción de las, 288; y homoplasia, 332; y teoría evolutiva, 325; y transformación, 285;
bacteriología, 366
bacteroides, 352
Baja California, 460, filogeografía de, 483
Baja California Sur, carpintero arlequín en, 487, 486
bancos de semillas, 70
barreras, biogeográfica, 469; ecológicas,

482; geográfica, 328; históricas, 482; reproductivas, 328, 333, 429, 430
barrido selectivo, 77
bases nucleotídicas, composición de las, 113
Bayes, teorema de, 31, 32, 141, 168
bayesiana, inferencia, 145, 470
Bering, estrecho de, 456, 457
Beringia, 456, 457
biogeografía, 88, 459
biología, de la conservación, 342; evolutiva, 428; molecular, 500; y marcadores moleculares, 542
BIOLOG™, 414
bio-medicina y marcadores moleculares, 542
bioremediación, 500
biparticiones, 149
Blast, 88, 406
BLOSUM, matrices, 120
BOOM, 513
bootstrap, 23, 97, 146; no paramétrico, 124, 145; paramétrico, 146; prueba estadística del, 472
Borrelia, recombinación en, 289
botánica e hibridación, 427, 434
Bradyrhizobium, 353, 354, 358; presencia de plásmidos en, 354; y clonalidad, 356
branch swapping. *Véase* ramas, intercambio de
branch-and-bound, búsqueda heurística, 472
Branta canadensis (ganso), 482
Bremer, medida de soporte de ramas de, 146
Bromeliaceae y pseudoviviparidad, 222
bromuro de etidio, 535, 550, 553
Brucella, 353
BSA (*Bovine Serum Albumin* o Albúmina Sérica Bovina), 530
bulbilo, 221, 222
Burkholderia, 354
búsqueda heurística exhaustiva, 472

C

Campylorhynchus brunneicapillus, 483
Coccidioides immitis, estructura panmíctica, 292
Cactaceae y pseudoviviparidad, 222
cactáceas, 506
cadena de Markov Monte-Carlo, 74
Callithrix jacchus (primate), 456
cambios homoplásicos, 445
Cameron, 28; y Kumar, 25
Campilobacter jejuni, 291; patrones poblacionales de, 151; y mutación, 339
Campylopterus curvipennis (colibrí fandanguero), 488; distribución disyunta, 488
Campylorhynchus brunneicapillus, secuencias de ADN mitocondrial en, 482
Canadá, 457; estudios filogeográficos, 480
captura citoplásmica, 435
caracteres, 131; ancestrales o primitivos, 94; apomórficos, 94; derivados, 94
carga genética, 288; moleculares, 541
Caribe, distribución de *Petrochelidon fulva*, 480
Carlia rubrigularis (lagartija), 457
carpintero arlequín, patrón filogeográfico del, 487
Centroamérica, distribución del carpintero arlequín, 486; estudio filogeográfico de *Wilsonia pusilla*, 481
Chao1, estimadores, 413

chapulines, hibridación en, 433
Chiapas, carpintero arlequín en, 487
Chihuahua, 483; estudio de *Meleagris gallopavo*, 481
chinches acuáticas, hibridación en, 433
Chlorospingus ophtalmicus, 488
cianobacteria. Véase *Synechococcus lividis*
ciclos, biogeoquímicos, 500; de vida, 240
ciencias forenses y marcadores moleculares, 542
cigoto, 184
citocromo-*b*, 450
cladograma, 90, 444, 477
clados anidados, 470, 487; análisis de, 57, 448, 453, 456, 476; distancia de, 477
Clevelandia ios, 454, 455
clina morfológica, 429, 435
clon, 243,
clonal growth, 222
clonalidad, 228, 242, 243; alta, 336; como estrategia, 243; y panmmixia, 290
clonas, 235, 413; librería de, 415
cloroplasto, 28, 49, 67, 78, 79, 543, 548; ADN del, 549; genoma citoplásmico del, 434; moléculas de, 544; rADN en, 544; y recombinación, 281
clustal X, 108; W, 109
coalescencia, 57, 59, 69, 73, 77, 78, 81, 82, 83, 101, 288, 335, 336, 447, 448; modelo de, 308; procesos de, 330, 447; simulaciones de, 308; tiempos de, 338, 474
coancestría materna, 474
Coccidioides immitis, estructura panmítica, 292; recombinación en, 292
codominante, 559
codones, 113, 117, 123; 294; uso preferencial de, 310
coeficientes, de coevolución, 88; de Dice, 166; de endogamia, 70, 78; de Jaccard, 167; de Nei y Li, 167; de selección, 80, 83; del rARN, 361
Cohan, 367, rutas evolutivas, 340
cohesividad, ecológica, 336; genética, 336
colibríes, morfología de los, 488
colonización, 82
Comité Internacional de Sistemática Bacteriana, 365
complejidad, hipótesis de la, 360
componentes, femenino, 173; masculino, 173
composición nucleotídica, 294; y recobinación, 293
comunidad, 393, 402, 406, 414, 461; bacteriana, 394; mésicas, 415
coníferas, 506
conjugación y recombinación, 285
consanguinidad, 171
conservación, 174; biología de la, 50, 80; y evidencia molecular de parentesco, 172
construcción filogenética, 42
contacto secundario, 429, 431
convergencias evolutivas, 331
conversión genética, 295
cópulas extra pareja, 171; y marcadores moleculares en aves, 171
Costa Rica, carpintero arlequín en, 486, 487
covarión, 122; modelo de, 121, 122; proceso de, 122; tipo, 123
cpADN, 435
Crassulaceae y pseudoviviparidad, 222
crasuláceas, 506
Cratogeomys (roedor de México), 458
crecimiento, 241; clonal, 219, 220; pobla

crecimiento *(continúa)*
cional, tasa de, 232, 242, tasa finita de, 241
cromátidas, 282, entrecruzamiento, 282; hermanas, 282;
cromatografía de gases, 398
cromosoma, 67, 68, 79, 354, 433; X, 78; Y, 79; Y, 453; y meiosis, 282
crossing-over. *Véase* entrecruzamiento
cruzas geitonogámicas, 230
Cryphonectria parasitica, recombinación en, 292
CTAB, 528
Cuaternario, eventos climáticos y geológicos del, 456
cuello de botella, 65, 82, 172, 188, 198, 482
cuitlacoche piquicurvo (*Toxostoma curvirostre)*, 475
Cylindropuntia y pseudoviviparidad, 222
Cynanthus doubledayi, 483, 484
Cynanthus latirostris (colibrí piquiancho), 483, 484
Cynanthus, 484; valor de F_{st}, 484; *sordidus*, 484

D

D de Tajima, 15
Darwin, 235; y Wallace, 11
datos morfológicos, 143
Datura stramonium; segregación mendeliana, 188; tasa promedio de polinización cruzada, 188
Dayhoff, matrices, 119, 120, 128; modelo de evolución de proteínas, 119
deme, 77
demografía, 84, 224, 446; histórica, 288, 84; y filogeografía, 446
Dendrobates (rana venenosa), divergencia genética, 458
densidad poblacional, 431
depresión por endogamia, 80, 197
deriva genética, 12, 14, 16, 50, 55, 63, 64, 71, 72, 79, 80, 82, 286, 300, 328, 329, 330, 331, 357, 447, 473
derivado o apomórfico, 94
dermatomicosis, 292
descomposición por biparticiones, 149
desequilibrio, 433; citonuclear, 431; de ligamiento, 42, 298, 339, 357, 369
detergentes no iónicos, 530
DGGE/TGGE, 402; geles, 407
dicógamas, 195, 184
dicotiledóneas, 220
diecia, definición, 203
diferenciación genética, 395; análisis de, 339; e isoenzimas, 544
diferencias macroevolutivas, 442
difusión, cámara de, 397
dinámica demográfica, 82
dinucleótidos, 517, 523, 529; frecuencias de, 293
dioicas, 184
dispersión, 173, 430, 443, 448; geográfica, 477; reducida, 224; tasas de, 173
distancia, evolutiva, 96; genética, 357, 369; ultramétrica, 96
distribución, agregada, 228; de Poisson, 64, 67, 69, 71; espacial, patrón de, 228; estable de edades, 230; gama, 121; lognormal, 41; *mismatch* multimodal, 474, 483
divergencia, genética (en tortugas gigantes de las Galápagos), 459; mitocon-

drial, 457; morfológica, 457; reciente, 469
diversidad, 406, 409; bacteriana, 393; clonal, 163; en ambientes acuáticos, 394; genética, 16, 184, 234, 235, 286, 288, 399; nucleotídica π, 14, 484; procarionte, mecanismos de control de la, 394
DL (desequilibrio de ligamento), 299
DMSO (dimetil sulfóxido), 530
dN/dS, 28, 30, 38, 40, 41
DNASP, 42
dNTP, 524, 529, 570
dominancia ecológica, 369
dominios, 104
Drosophila, e isoenzimas, 544; *melanogaster*, 19, 477; *simulans*, 19, 39; *yakuba*, 19, 39
dS/dN, 33
duplicación de genes, 283

E

eBURST, algoritmo de, 151
ecología, 365, 396; de poblaciones, 224; microbiana, 404, 412, 415; y marcadores moleculares, 542
ecotipo, 367, 368
ecotonos, 431
Ecuador, distribución de *Petrochelidon fulva*, 480
edades, estructura estable, 232, 241
EDTA, 501
eigenvectores, 231, 232
elasticidad, 242; análisis de, 224, 242
Elephas maximus (elefante asiático), 459
elementos móviles, conjugación de, 294
elisión, método de, 124
EMBL, 37
enantiostílica, 195
endogamia, 65, 184, 185, 187, 194; biparental, 192, 198, 199; coeficiente de, 70, 78, 187; depresión por, 80, 197; estimación de la, 183; tasa de, 199;
endonucleasas, 552
enjambres híbridos, 427, 434
enterocinasa, 104
entidades, biológicas, 364; evolutivas, 337
entrecruzamiento, 282, 333, 337; desigual, 283; tasa de, 197
enzimas, estudios de variación genética con, 467; de restricción, 143
epidemias, clonas hipervirulentas en, 292
epidemiología, 82, 163
equitatividad (evenness), 412
Escherichia coli, 38, 285, 291, 340, 361, 367, 368, 405; ADN en, 399; genoma de, 291, y transformación, 285
esfíngidos, 188
especiación, 101, 326, 329, 340, 367, 467; parapátrica, 333; simpátrica, 333
especies, abundancia de, 411; alopátricas, 459; asexuales, 367; autógamas, 229 bacterianas, 357; clonales, 228, 234; concepto biológico de, 450; concepto arbitrario de, 337; concepto cohesivo de, 330, 334; concepto ecológico de, 328, 362, concepto evolutivo de, 327, 363; concepto filogenético de, 329, 334, 337, 450; concepto formal de, 325; concepto genético de, 329; concepto morfológico de, 362; concepto tanoxómico de, 362; definición
especies *(continúa)*
biológica de, 290, 328, 333, 366, 367,

especies *(continúa)*
414, 450; definición pragmática, 366; descripción de, 326; distribución de, 415; estatus de, 320; marinas, estructuración poblacional de las, 454; microbianas, 414; raras, 413; nuevas, 395; número de, 412; problemas con el concepto de, 336; riqueza de, 410; simpátricas, 459;
espectrometría de masas, 398
esporas asexuales, 292
estadística filogeográfica, 448
estadísticos de F, modelos matemáticos, 305, 331, 447, 473
estado ancestral o plesiomórfico, 94
Estados Unidos de América, distribución de *Petrochelidon fulva* en los, 480; distribución del carpintero arlequín en los, 486; estudio de *Wilsonia pusilla* en los, 481; estudios de filogeografía comparativa, 482; estudios filogeográficos, 480; trabajos filogeográficos en los, 450
estándar reconocido, 366
estimadores no paramétricos, 412, 413
estocasticidad demográfica, 69
estrategia mixta, 195
estructura, filogeográfica, 453, 480; genealógica de las poblaciones, 470; genética, 82, 432, 434, 467; geográfica, 430; evolutiva, 342; poblacional, 72; terciaria, 406
estudios, de diversidad y marcadores moleculares, 542; filobiogeográficos, 435, 458; filogeográficos, 435, 448; forenses, 163
etanol, 500, 502
Eucalyptus, brownii, 435; *populnea*, 435
eucariontes, 401, 500; células, 499; intercambio genético, 358; evolución del sexo, 183, recombinación en, 286, 311
Eucarya, 401
Eucyclogobius newberryi, 455
Euglena, 542
Euler-Lotka, ecuación de, 230
Europa, trabajos filogeográficos en, 457
eventos, filogeográficos, 448; vicariantes, 457
evolución, 396; bacteriana, 365, 366; celular, 359; concertada, 283; experimental, 395; genómica, 517; mínima, 129, 130, 131; molecular, 12, 13, 42, 83, 125, 183, 517; tasa de, 444; y marcadores moleculares, 542
exclusión, 168, 170
éxito reproductivo, 68, 69, 70
exones, 543
exploraciones heurísticas, 134
extinción, 49, 80

F

factores ecológicos, 430
fagos, 359
falange, 217
familias multigénicas, 283, 549
fecundación, cruzada, 184, 188, 193, 197; mixta, 198; tasas de, 185, 190, 191, 192, 194, 199; vegetativa, 241
fecundidad, 220
Felsenstein, Joe, 42; método de, 139; modelo de, 115
fenotipo, 358; fenotipos/genotipos, 11
Ferocactus robustus, distribución de ramets en, 228
fijación, índices de, 473

filogenética tradicional (macroevolución), 469
filogenia, 58, 87, 89, 99, 101, 122, 127, 147, 295, 334, 449, 541; comparada, 483; con PCRs, 521; método de estimación de, 127
filogeografía, 78, 82, , 88, 441, 442, 447, 449, 454, 461, 467; comparada, 459-461; intraespecífica, 453; métodos y análisis de la, 470-479; perspectivas, 450; publicaciones científicas sobre, 467
filograma, 444
filogrupos, 455
filopatría, 78
fingerprinting, 163
F_{IS}, 569
FISH (acción fluorescente *in situ*), 399, 408, 415
Fisher, R. Véase máxima verosimilitud
F_{IT}, 569
Fitch-Margoliash, método, 129; algoritmo, 130
flores hermafroditas, 184
flores perfectas, 183
fluctuaciones climáticas, 456
flujo génico, 42, 49, 51, 55, 57, 58, 81, 82, 173, 329, 330, 331, 333, 334, 363, 395, 434, 447, 452, 455, 469, 473, 482, 484 monitoreo del, 548; restringido, 479
formamida, 530
formato Newick o New Hampshire, 90
fosfolípidos, 397
fotosíntesis, 311
frecuencias alélicas, 15, 52, 167
frijol, evolución en Mesoamérica y Sudamérica, 368; criollo, 369
frutos abortados o plantlets, 222
F_{st}, 52, 55, 59, 72, 73, 74, 75, 78, 483
fuerzas, ecológicas, 330, 334, 336; evolutivas, 11; genéticas, 330, 336
Fu-Li, 14, 15
función exponencial negativa, 410
fundador, efectos del, 188, 198

G

G/N, 235
gama discreto (método), 116
gametofito, 222
ganso, 482
gap, 106
gases, cromatografía de, 398
geitonogamia, 173
geles desnaturalizantes de gradiente (DGGE), 399
GenBank, 37, 166, 406
GenCompress, programa de cómputo, 127
genealogías, 427, 441, 442
generaciones discretas, 70
género, 367
genes, 100; copias del, 407; del núcleo y tiempo de coalescencia, 336; esenciales, 360; intercambio genético, 368; mitocondriales, 450, 458, 460; mitocondriales ND2, 475; ortólogos, 100, 361; parálogos, 100
genética de poblaciones (microevolución), 78, 167, 234, 281, 351, 443, 448, 467, 469, 472, 517, 556
genética, 430
genets, 173, 216, 218, 223, 224, 230, 235; en *Opuntia rastrera*, 227
genoma, 13, 38, 49, 98, 166, 183, 286,

genoma *(continúa)*
293, 294, 361, 407, 521, 544; celulares, 541; citoplásmico, 435; de *E. coli*, 291; evolución de los, 125; de una bacteria, 150, 364; de un metanógeno, 150; completo, 354, 359; humano, 165; recipientes, 360; mitocondrial (mtADN), 543; nuclear eucarionte, 567; recombinación, 281, 282, 293; regulación del tamaño del, 281; ribosomal, 541
genómica, potencial de la, 88
genotipo-ambiente, asociaciones, 434
genotipos, 241; heterócigos, 545; homócigos, 545; y recombinación, 286
Geochelone nigra (tortugas gigantes de las Galápagos), 458
GeoDis, programa, 58, 477
geografía, 448
Geomys pinetis, 449
Gesneriaceae y pseudoviviparidad, 222
Gibberella fujikuroi, genealogías generadas en, 335
gibón, 93
ginodioicas, 195, 198
GIPS, 513
glaciaciones, 429
glnII, gen, 358
globinas, 34
gold standard. *Véase* estándar reconocido, microsatelitales de, 445; región control de, 445
Gossypium (algodón), 428
gramíneas, 221
grillos, hibridación en, 433
grupos híbridos y flujo génico, 333
grupos naturales, 331
Gryllus firmus, 430; *pennsylvanicus*, 430
Gst, 53
guerrilla, 217

H

hábitat, factores del, 334
Haemophilus influenzae y transformación, 285
haplotipos, 57, 337; 453, 460, 472; red de, 477
Hardy-Weinberg, 64; equilibrio de, 187, 546, 547, 552; modelo de, 189; población bajo equilibrio según, 168
Helianthus (girasol), 428
Helicobacter pylori, 38; y recombinación, 289
herbívoros, 228; y pérdida de biomasa, 227
hercogamia, 184, 204
heredabilidad, cálculo de, 172
herencia, 442; estrictamente materna, 442; materna, 78
hermafroditas, 184, 195, 215; autocompatibles, 195
herpetofauna, 454
herramientas, filogeográficas, 456; moleculares, 330, 415
heterócigo, 568; e isoenzimas, 544; individuos, 545
heterocigocis, 63, 64, 65, 69, 234
heteroduplex, hebras, 285
heterostílicas, 195; plantas, 204
hibridación, 423, 424; ADN-ADN, 335, 338; natural, 427; superioridad restringida de la, 431; y RFLPs, 552; zonas de, 429-434;
híbrido, 430-432; ancestro, 428

hipótesis de la complejidad, 360
historia genealógica, 447
historias de vida, 51, 216, 230; a nivel de ramets, 224
hitchhiking. *Véase* aventón
HLA (sistema de detección),166
Homo sapiens, 453; migraciones, 453; origen africano del, 453
homócigo, 171, 568; individuos, 545
homología, de las bandas, desventajas, 569; morfológica, 541
homoplasia, 81, 288, 311, 445
hongos, 398; concepto biológico de, 337; dinámica evolutiva, 310, 325; estudio de la recombinación en, 292; estudios filogenéticos de los, 556; intrones y, 543; mecanismos de reproducción, 288; recombinación, 310; y homoplasia, 332
hormigas, análisis de paternidad, 164; hibridación en, 433
Hudson-Kreitman-Aguadè, 14, 16, 42
huellas digitales, 164

I

IA de Maynard-Smith *et al.,* 299
incompatibilidad, sistema de, 194, 204
incongruencias filogenéticas, 288
índice, D, 412; de asociación, 299
individuos, 442
Indonesia, 459
infecciones periodontales, 290
inferencia filogenética, 311; efectos de la recombinación en, 293
insectos, 454
intercambio genético, 359, 360, 361
introgresión, 548
intrones, 543
Iris, 434
islas, de patogenicidad, 360; genómicas, 358; oceánicas, estudios de genética y conservación en, 458; simbiótica, 354.
isoenzimas, 433, 547; isoenzimas/aloenzimas, 547; modelo de, 174, 189, 192
ISSR (*intersequence simple repeats*), 239, 550-552, 567, 569; marcadores, 165, 198, 520
iteróparo, genet, 223
ITS1 e ITS2, 544

J

Jaccard, coeficiente de, 167
Juncaceae y pseudoviviparidad, 222
Jukes-Cantor, 22, 40

K

Ka/Ks, 39
Kimura, 13, 39, 20, 25, 40; parámetros de, 24, 25, 304
Kosakovski, Sergei, 41
Kumar, Sudhir, 41; *et al.*, 28

L

laboratorios naturales, como zonas de hibridación, 428
LAMARC, 75
Lampornis amethystinus (colibrí garganta amatista), 483, 487, 484; *amethystinus*,

Lampornis amethystinus (continúa) 486; *margaritae*, 486; *salvini*, 485
landraces. Véase frijoles criollos
Larrea tridentata, longevidad de clones de, 217
leguminosas, 351, 352, 355, 506
leopardo, 455
Lepus americanus, 457
Letharia (hongo liquenizado), recombinación en, 292
levaduras, tamaño del genoma mitocondrial en, 543
ligamento citonuclear, desequilibrio de, 434; desequilibrio del, 433, 434
likelihood ratio o LOD. *Véase* máxima verosimilitud
Liliaceae y pseudoviviparidad, 222
linaje, alopátricos, 469; chimpancé-gorila-humano y coalescencia, 336; evolutivo, 364;
simpátricos, 469
Linneo, concepto de especie de, 327
lisozima C, 33
Lithocarpus, 435
Li-Wu-Luo, 25, 27
loci, dominantes, 547; polimórficos, 234
LOD *(likelihood ratio)*, 170
Lotus corniculatus, 355

M

M (= flujo génico), 53
Macaca mulatta, 172; áreas de endemismo, 460
macacos, identificación de paternidad, 172
macroevolución, 442, 468
macroorganismos, 343
magueyes, 221
maíz, en Mesoamérica, 369
Malasia, 459
mamíferos, 450, 452, 461; de Amazonía oriental, 457; de talla grande, 455; estructura filogeográfica, 460; hibridación en, 433; marinos, 452; voladores, 452
Mammillaria, módulo apropiado para su descripción, 218
marcadores, citogenéticos, 428; citoplásmicos, 438; dominantes, 542, 546, 568; electroforesis multilocus de isoenzimas, 369; haploides, 82; genéticos, 434, 567; moleculares, 163, 171, 174, 184, 195, 332; microsatélites, 55, 74, 165, 194; nucleares, 434, 450; químicos, 428; y patrones conductuales no observados directamente, 171. Véase RAPDs, AFLPs, ISSRs
marcaje-liberación-recaptura (MRR), estadísticos de, 412
Markov, cadenas de, 293
mARN, 359, 514, 559
material genético y recombinación, 281
MATLAB, programa, 241
matriz, de sensibilidad, 242; de transición, 240, 242
máxima parsimonia, 32, 33
máxima verosimilitud, 20, 28, 33, 41, 43, 58, 73-76, 168, 169, 170; métodos de estimación, 305
McDonald-Kreitman, 14, 18, 19
mecanismos, sexuales, 356; parasexuales, 356
medicina forense, 517

MEGA 2.1 (programa para árboles de distancia), 471
meiosis, 282, 300
meiosporas, 337
Melanerpes formicivorus (carpintero arlequín), 483, 486, 487
Meleagris gallipavo (pavo), 481; *gallopavo*, 481; *intermedia*, 481; *merriami*, 481; *mexicana*, 481
Melospiza melodia, 480, 482
Mendel, leyes de, 472
meristemo axilar, 222
Mesorhizobium, 358; presencia de plásmidos en, 354
metabolismo, 414
metapoblación, 50, 77, 80
Methylobacterium nodulans, 354
metilobacteria, 353
métodos, bayesianos, 305, 309; de asignación fraccionada y categórica, 167; de compatibilidad, 298; de coalescencia, 76; de distancia, 295; de exclusión, 167; de potencia, 232, 241; de reconstrucción genotípica, 167; demográficos, 66; directos, 50; directos e indirectos, 66; filogenéticos, 295, 298, filogeográficos, 82; genealógicos, 57; indirectos, 51, 71, 81; no paramétricos, 300
Metropolis-Hastings, algoritmo de, 75; aproximación de, 74; muestreo de genealogías, 75
México, distribución del carpintero arlequín en, 487 y PCR, 523; estudio de *Wilsonia pusilla* en, 481; historia biogeográfica de las aves, 483
MHC (sistema de detección), 166
microarrays. *Véase* microarreglos.
microarreglos, 559, 560, 561, 562,
microbios, 407, 413, 415
microevolución, 442, 468
microorganismos, 393, 396, formación de nuevas especies, 333; reproducción de los, 288; y concepto genético de especie, 329
microsatélites, 433, 450, 455, 521, 549; codominantes, 548
microscopio óptico, 408
Microtus oeconomus, 457
Michaelis-Menten, ecuación de, 410
migración
migración, 51, 66, 73, 81, 82, 225, 448; tasa de, 456, 473
milpa, 369
minimum evolution, 470
mismatch, gráficas de, 475
mitocondria, 49, 67, 68, 78, 450; genoma citoplásmico de las, 434; moléculas de, 544; rADN en, 544; y recombinación, 281
mitosis, 282
ML. *Véase* máxima verosimilitud
modelo, de islas, 55, 72; de islas infinitas, modelo *(continúa)* 51; de islas de Wright, 73; demográfico, 228; de Wright-Fisher, 73; de sitios infinitos, 17, 74; mutacional SSM, 549; neutro, 13, 17; poblacional, 240; Poisson normal, 411
módulo, 218, 220
moluscos, hibridación en, 433
monocárpica, roseta, 223
monocotiledóneas, 220
monoecia, definición, 203
monofilia, 362, 363
monogamia, 172

monoicas, 184
monoicismo, 193
Monte Carlo, simulaciones de, 299
Montes Urales, 457
morfología, 480, 542
mortalidad, 225
mosaico, concepto de, 430; zonas en, 433
mRNA, 552
mtADN, 444-447, 469, 453, 455, 456, 459; en individuos de Brasil y Portugal, 453; en animales, 453, 543
muestra, tamaño de, 410, 413
muestreo, calidad del, 409; esfuerzo del, 409; estrategias de, 415
murciélagos, 460
Muse, Spencer, 41
Museo de Zoología, Colección de Tejidos de Aves del, 490
mutación, 12, 16, 17, 64, 80, 81, 83, 165, 166, 171, 286, 305, 329, 330, 338, 339, 340, 357, 473; adaptativa, 340; como efecto deletéreo, 197, 286; contribución a la variación, 291; en grandes camadas, 172; simple, 166; tasa de, 305, 308

N

NADH (alcohol deshidrogenasa), 34, 477
nADN, 543
NCA, 449, 459
NDH, genes mitocondriales, 450, 458
Ne, 64-72, 74, 75-83, 308
néctar, 184
Nei y Li, coeficiente de, 167; distancias genéticas de, 240
Neigel, análisis de, 470
neighbor joining, 32, 470, 472
Nei-Gojobori, 20, 24, 25, 27, 32, 42
Neisseria, ausencia de clonalidad, 356; *gonhorreae* y transformación, 285; *meningitides*, 38, 291, 340; recombinación en, 290; y mutación, 339
nested clade analysis, 57
nicho, 328, 358; ecológico, 340, 367
nitrógeno, 354, 507; fijación de, 354
Nm, 53, 55, 474
nod, genes, 354
nódulos, 351, 352
nopal rastrero, tallas de los ramets de, 217
nucleasas, 501
núcleo, rADN en, 544
nucleótido, 13, 41; extras, 567; secuencias de, 444
Nueva Zelanda, 355
número efectivo de migrantes, 74

O

Odontesthes perugiae, origen, 459
Ohta y Kimura, modelo mutacional SSM, 549
oligonucleótidos, 400, 408, 521, 523, 530, 560
Oncorhynchus gorbuscha (salmón rosado), 459
operones de rARN, 538, 407
Opuntia, módulo apropiado para su descripción, 218; *rastrera*, ciclo de vida complejo, 227; y pseudoviviparidad, 222
origen de las especies, El, 325
ortólogos (familias de genes), 293; uso en métodos empíricos, 294
ostrácodos, recombinación, 310

OTUs, 409, 413, 414
Ovis canadensis, 512
Oxalidaceae y pseudoviviparidad, 222

P

Packera, 435
Paeonia, 428
Pamilo-Bianchi-Li, 25, 27, 28, 38
Pan troglodytes (chimpancé), 93, 455; *verus*, 455
panmíctica, población, 65, 72, 77; estructura aparentemente, 292
panmixia, 64, 68, 229
Panthera onca (jaguar), estructura geográfica, 457; *pardus* (leopardo), 455
parálogos (familias de genes), 293
parámetro maltusiano, 230
parentesco, 172
parsimonia, árboles de, 472
paseriformes, análisis de parentesco, 172; paternidad múltiple en, 172
Passerina iliaca (gorrión), 482
paternidad, 166, 167, 169 (nota 1), 174; múltiple, 172; análisis de 163, 164, 165, 170, 174
patogenicidad, islas de, 358
patógeno, 340, 358
patrones filogeográficos, 441, 470, 482, 483; del carpintero arlequín, 487; geográficos, 459; regionales filogenéticos, 459
PAUP, programa para árboles de distancia, 471
PCR (*Polymerase Chain Reaction*), 399, 400-402, 404, 406, 407, 414, 517, 520, 521, 524, 527, 530, 531, 532, 534, 555; y marcadores dominantes, 192; y RFLPs, 452
peces, 450, 452; de agua dulce, 450, 482; de agua salada, 482; hibridación en, 433; marinos, 452; marinos simpátricos de California, 455
pedigree, 186. *Véase* árbol genealógico
pequeños mamíferos, 454; no voladores neotropicales, 460
perfil de ADN, 164
Peromyscus furvus, diferenciación genética, 458; *maniculatus*, 454; *polionotus*, 454
Perú, distribución de *Petrochelidon fulva*, 480
Petri, cajas de, 396
Petrochelidon fulva, 480; *citata*, 481; *pallida*, 480
pez de desierto *(Gila atraria)*, patrones filogeográficos, 454
PFLA (*Phospholipid Fatty Acid Analysis*), 397-398
Phacelia dubia y fecundación cruzada mixta, 198
Phaseolus vulgaris, 368; domesticación, 368
phyla, 467
Phylloscopus trochiloides, 455
piedras de paso, 73
pino-encino, y distribución del carpintero arlequín, 486
Pipilo fuscus, 483
plantas, análisis de paternidad en, 173; clonales, 217, 219, 228, 243; demografía de las, 224; e intrones, 543; estudios filogenéticos en, 456; hibridación en, 432; nuevas especies de, 333; suculentas, 509; superiores, tamaño del

genoma mitocondrial en, 543
plásmidos, 354, 357, 358, 359; conjugación de, 294; simbióticos, 355, 358, 359
Pleistoceno, 456, 460
Pleorotus ostreatus (hongo), especies de, 333
Plioceno, 456
Poaceae y pseudoviviparidad, 222
población, 427, 442, 443, 489; clonal, 290, 356; definición de, 219, 400; ideal, 73, 77; natural, 446, 473; panmíctica, 308; pequeña, 474; tamaño efectivo de la, 172, 308;
Poisson, distribución de, 64, 67, 69, 71
polen, 71; flujo de, 174
poliacrilamida, gel de, 545
policárpicas, especies, 223
poligamia, 172
polimerasa, 523, 529, 557; Reacción en Cadena de la, 517
polimorfismo, 16, 548, ancestral, 335; estilar, definición, 204; interespecífico, 335
polinización cruzada, tasas de, 188
polinizadores, 198
Polioptila californica, 483; *melanura*, 475, 483
Polygonaceae y pseudoviviparidad, 222
Popolus, 434
Porphyromonas gingivalis, recombinación en, 290
power method. Véase método de potencia
primers, 551, 567, 569
probabilidad, condicionada, 169 (nota 1); posterior, 168
procariontes, 351, 401, 499, 500; aplicación del concepto de especie a los, 338; concepto biológico de, 339; intercambio genético en, 286; mecanismos de recombinación en, 285; métodos para la delimitación de especies en, 332; y concepto genético de especie, 329
procesos, demográficos, 226; filogeográficos, 448; microevolutivos, 442
Pronolagus rupestris (conejo rojo de Smith), 454
propagación vegetativa, 221, 228, 229, 242
protándrica, 204
proteínas, 501; análisis de, 543; síntesis de, 360
proteobacterias, 352, 354
Proteus vulgaris, 361
prueba de Tajima, 14
pseudo-máxima verosimilitud, 76
Pseudomonas stutzeri, variación genética de, 289
pseudomonas, alta clonalidad, 356
pseudoviviparidad, 222

R

rADN 16S, 360, 361, 401, 406, 407, 544; coevolución del, 361
Ralstonia, 354
ramas, intercambio de, 138, 141
ramets, 173, 216, 218, 220, 224, 225, 230, 234, 235; de *Opuntia rastrera*, 227. *Véase* también clon
Rana camerani, 456; *holtzi*, 456; *macrocnemis*, 456
ranas de Anatolia (Turquía), 456
RAPDs (marcadores), 165, 189, 194, 198, 239, 400, 433, 520, 545, 547, 548, 550;

estudios de variación genética con, 467
rarefacción, 410; curva de, 410
rARN, 399, 401, 402, 406, 407, 408; rARN (SSU), 401
reciclaje de nutrientes, 500
recolonización, 80
recombinación, 42, 288, 289, 310, 330, 339, 340, 355, 357, 358, 360, 444, 473; como fenómeno excepcional, 294; definición, 281; en *Pseudomonas stutzeri,* 289; frecuencia, 356, en bacterias, 357; homóloga, 355, 356; ilegítima, 360; indetectable, 295; intraespecífica, 340; meiótica, 165; mecanismos de la, 286; métodos de detección de, 294; métodos de similitud y, 295; métodos empíricos para detectar, 293; métodos no paramétricos de; 294, 300; por *crossing-over*, 282; tasa de, 304, 306, 311; y cadenas de Markov, 293; y variación, 291
reconstrucción filogenética, 32, 33, 335; genotípica parental, 171
red (o árbol) de haplotipos, 477
región control, 446. 450; I y II, 475; en vertebrados y equinodermos, 444
regiones microsatélite, 567
reinos, 467
relaciones filogenéticas, 330
reloj molecular, 288, 445
reproducción sexual, 29, 228, 229, 242, 333; reproducción sexual, concepto biológico de 337; y eucariontes, 286
reptiles, 450; diferenciación filogeográfica, 452; estructura filogeográfica, 460; patrones filogeográficos, 461
retrocruzas, 427
retrovirus, material genético de los, 499
RFLPs (marcadores), 165; del mtADN; 433; del nrADN, 433
RFLPs, 547, 548, 552; estudios de variación genética con, 467; técnica de, 553
Rhizobiales, 352
Rhizobium, 352, 354, 358; *etli* bv. *phaseoli*, 368; *etli* bv *mimosae*, 368; *gallicum*, 368; *gardinii*, 368; *leguminosarum* bv. *phaseoli*, recombinación de, 290, 368; y clonalidad, 356
Rhodopseudomonas, 353
ribosomas, 401
riqueza biológica, 342
rizobios, 351, 353
rizósfera, 352
roedores, 449
rotíferos, recombinación, 310
rrl, gen, 358
rrs, gen, 352, 358, 365
Rusia, 457
ruta, análisis de, 241, 242

S

Saccharomyces cerevisae, 361
SAGE (*serial analysis of gene expression*), 166
salmón rosado, linajes de, 459
Salmonella, alta clonalidad, 356; enterica, 291; *typhimurium*, 361
San Miguel Acuexcomac, variedades de frijol en, 369
Sanger, Fred, reacción de, 559
sapos del género *Bombina*, hibridación en, 433

saprófagos, 352
Saxifragaceae y pseudoviviparidad, 222
Sclerotinia y teoría de la coalescencia, 337
SDS, 528
secuencia 16S rARN, 332; y filogenia, 332
secuencias, 557; alineación de, 471; de transformación, 285; nucleotídicas, 352
selección, 12, 16, 55, 330, 357, 408, 429-432; adaptativa, 13; coeficientes de, 241; ecológica, 367; exógena, 434; intensidad de la, 429; natural, 11, 20, 79, 83, 188, 286, 288, 295, 300, 311, 325, 327, 329, 331, 367, 433, 457; negativa, 12, 15, 33, 38, 83; positiva, 12-16, 20, 23, 29, 32, 33, 34, 38, 38; purificadora, 33, 39; periódica, 340
selective sweep, 77
semillas, 71
sensibilidad, 232; matriz de, 242
sensibilidades, 241
serial analysis of gene expression. Véase SAGE
sexos, evolución de los, 183; separados, 65
Shannon-Weaver, índice de, 412
Shigella, 368
Siberia, distribución de *Phylloscopus trochiloides*, 455
SIDA, 83
simbiosis, 352, 360; islas de, 358, mutualista, 352
simbiosoma, 352
similitud, 362; ADN:ADN, 366
Simpson, índice de diversidad de, 235, 412
simulaciones de Monte Carlo, 299
single nucleotide polymorphism. *Véase* SNPs
Sinorhizobium, 352, 354
sistema, de apareamiento, 70, 183; de reproducción, 183, 545; dominante recesivo, 557
sistemática, bacteriana, 366; e hibridación, 434
sitios infinitos, 80, 308
sliding windows. *Véase* ventanas deslizantes
SNPs, 166
sobrevivencia, 220
Sonora, 483; desierto de, 460; estudio de *Meleagris gallopavo* en, 481
sosa, 500
southern blots, y RFLPs, 552
Spizella passerina, 482
split decomposition, método, 310
SSCP, 520, 521
SSRs, 548, 567
SSU, 401
Staphylococcus aureus, recombinación en, 291; y mutación, 339
stepping stone, 73
stepwise addition, búsqueda exhaustive, 472
step-wise model, 55
Streptococcus pneumoniae, 291; recombinación en, 291; y transformación, 285; y mutación, 339
Streptoprocne zonaris, 488
sucesión, 395
suelo, 398; diversidad genómica en, 394; y comunidades microbianas, 398
Sulawesi (Indonesia), 460
sustitución, 444; de nucleótidos, 446; sinónima, 295
Synechococcus lividis (cianobacteria), 332

T

programa TCS, 477
Toxostoma curvirostre (Cuitlacoche piquicurvo), 475, 483; *lecontei arenícola*, 481; *lecontei lecontei*, 481
Tajima, 14, 15, 16; McDonald y Kreitman, 42
tamaño, del vecindario, 70; efectivo, 13, 15, 59, 63-84, 455, 546; poblacional, 63, 73, 82, 474
taq, 524
tARN, 514
tasa, de especiación, 395; de evolución, 335; de extinción, 395; de fijación; 83, de migración, 51-53, 74; de mutación, 13, 14, 19, 71, 300, 305, 549; de sustitución, 33, 34, 76, 80, 288; de recombinación, 304
taxonomía e hibridación, 434
Templeton, A., 479; análisis de clados anidados, 477
tensión, zona de, 430
teorema de Bayes, 31, 32, 168
teoría, de coalescencia, 446-449; evolutiva, 363; neutra, 12, 13, 17, 83; y *Sclerotinia*, 337
termociclador, 528, 524
TFPGA, 570
TGGE, 406
Thermus aquaticus, 517
Thymus vulgaris y fecundación cruzada mixta, 198
tiempo, al ancestro común más cercano, 308; generacional, 66
tm (tasa de fecundación cruzada multilocus), 194, 195
TMRCA, 308
tokogenia, 334
tortugas, diferenciación filogeográfica, 452; gigantes de las Galápagos, 458; marinas (análisis de paternidad), 164
Toxostoma lecontei, distribución filogeográfica, 481; estructura filogeográfica, 481; interrupción filogeográfica en, 482
transducción, 285, 294
transferencia horizontal, 291, 294, 356; y evolución celular, 359
transferencia lateral, 355, 357, 358, 360, 361
transformación (mecanismo de recombinación), 285
transición, matriz de, 242
transposones, 359
TRF, 403
TRFLPs, 404
Trychophyton rubrum, 292
Turdus infuscatus, 488

U

úlceras estomacales, 289
unidades, evolutivamente significativas, 331, 341, 342, filogeográficas, 456
Unidades Significativamente Evolutivas, 458
UPGMA, 472
uptake sequences. *Véase* secuencias de transformación
Ursus arctos (oso café), extinción de, 457

V

valor reproductivo, 241
varianza, 299; genética, 286
vecindario, 70, 73, 81
ventanas, deslizantes, 295; de hibridación, 428
verosimilitud, 170
VIH-1, virus, 83
virus, y bacterias, parasitismo entre, 404; y transducción, 285
variación, genética, 15, 17, 79, 80, 340, 369, 432, 548; geográfica, 428; haplotípica, 448; neutral, 81; molecular, 541; morfológica, 541
vertebrados, 450; mexicanos y filogeografía, 461
Viola pubescens, 551
Vibrio cholerae, ausencia de recombinación en, 310

W

W, 29, 30 ,41
Wilsonia pusilla (chipe), 455, 481; estudio filogeográfico de, 481
Wright, 473, 569; F de, 474
Wright, Sewall, 51, 52, 53, 63, 64, 65, 187, 447. *Véase* coeficiente de endogamia

X

xenógamas, especies, 229
xenogamia, 173

Y

Yucatán, distribución de *Petrochelidon fulva*, 480

Z

zona híbrida, 333
zoólogos e hibridación natural, 428

Ecología molecular, compilado por Luis E. Eguiarte, Valeria Souza y Xitlali Aguirre se terminó de imprimir y encuadernar en los talleres de Impresora y Encuadernadora Progreso, S.A. de C.V. (IEPSA), Calzada de San Lorenzo 244, 09830, México, D.F. durante el mes de septiembre de 2007

Se tiraron 600 ejemplares